Mehmet Eren NALİCİ

M.Sc. Thesis

AGU 2024

NAVIGATING BIST100 INVESTMENTS THROUGH SYMBOLIC AGGREGATE APPROXIMATION CLUSTERING: INSIGHTS FOR INVESTORS

M.Sc. THESIS

SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF ABDULLAH GUL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Mehmet Eren NALİCİ

May 2024

NAVIGATING BIST100 INVESTMENTS THROUGH SYMBOLIC AGGREGATE APPROXIMATION CLUSTERING: INSIGHTS FOR INVESTORS

A THESIS
SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Mehmet Eren NALİCİ
May 2024

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Mehmet Eren NALİCİ
Signature:

REGULATORY COMPLIANCE

M.Sc. thesis titled “Navigating BIST100 Investments Through Symbolic Aggregate Approximation Clustering: Insights for Investors” has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor	Co-Advisor
Mehmet Eren NALİCİ	Assoc. Prof. Dr. Ramazan ÜNLÜ	Dr. İsmet SÖYLEMEZ
Signature	Signature	Signature

Head of the Industrial Engineering Program

Assoc. Prof. Dr. Ramazan ÜNLÜ

Signature

ACCEPTANCE AND APPROVAL

M.Sc. thesis titled “Navigating BIST100 Investments Through Symbolic Aggregate Approximation Clustering: Insights for Investors” and prepared by Mehmet Eren Nalici has been accepted by the jury in the Industrial Engineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

…….. /…….. / ……..

(Thesis Defense Exam Date)

JURY:

Advisor: (Assoc.Prof. Dr. Ramazan ÜNLÜ) ……...…….….....

Member : (Assist. Prof. Dr. Ahmet ERDOĞAN) …...…….…….....

Member : (Assist. Prof. Dr. Betül ÇOBAN) …...……..……. .....

APPROVAL:

The acceptance of this M.Sc./Ph.D. (choose one) thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated …. /….. / …….. and numbered.…….…….

…….. /…….. / ……..

(Date)

Graduate School Dean
Prof. İrfan ALAN

ABSTRACT

Navigating BIST100 Investments Through Symbolic Aggregate Approximation Clustering: Insights for Investors

Mehmet Eren NALİCİ
MSc. in Industrial Engineering Program
Advisor: Assoc. Prof. Dr. Ramazan ÜNLÜ
Co-advisor: Dr. İsmet SÖYLEMEZ

May 2024

Market stakeholders, including traders and investors, strive to forecast stock market returns for informed decision-making. Computational finance employs various tools such as machine learning techniques to analyse extensive financial datasets to provide predictive insights for investors. Among all those techniques, clustering is one of the most well-known and used machine learning methods to reveal hidden patterns from unlabelled data. This study aims to help investors make more robust decisions by autonomously identifying companies that may exhibit similar price movements. In our study, with the model developed based on the Symbolic Aggregate Approximation (SAX) method, BIST100 companies are divided into clusters of various numbers and various scenarios are developed for investors from different perspectives such as risk minimization and strategic investment. The SAX clustering method is employed for analysing share movements. Moreover, dendrogram tree graph is used to analyse the clustering of different SAX combinations.

Keywords: Machine Learning, Symbolic Aggregate Approximation (SAX), BIST100, Stock Market

ÖZET

Sembolik Toplam Yaklaşım Kümelemesi Yoluyla BIST100 Yatırımlarında Yön Bulma: Yatırımcılara Yönelik Bilgiler

Mehmet Eren NALİCİ
Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans
Tez Danışmanı: Doç. Dr. Ramazan ÜNLÜ
İkinci Tez Danışmanı: Dr. İsmet SÖYLEMEZ

Mayıs 2024

Ticaret ile uğraşan kişiler ve yatırımcılar da dahil olmak üzere piyasa paydaşları, bilinçli karar verme amacıyla borsa getirilerini tahmin etmeye çalışmaktadır. Hesaplamalı finans, yatırımcılara öngörücü bilgiler sağlamak amacıyla kapsamlı finansal veri kümelerini analiz etmek için makine öğrenimi teknikleri gibi çeşitli araçlar kullanır. Tüm bu teknikler arasında kümeleme, etiketlenmemiş verilerden gizli kalıpları ortaya çıkarmak için en iyi bilinen ve kullanılan makine öğrenmesi yöntemlerinden biridir. Bu çalışma, benzer fiyat hareketleri sergileyebilecek şirketleri otonom olarak tespit ederek yatırımcıların daha sağlıklı kararlar almasına yardımcı olmayı amaçlamaktadır. Çalışmamızda Sembolik Toplam Yaklaşım (SAX) yöntemi esas alınarak geliştirilen model ile BIST100 şirketleri çeşitli sayıdaki kümelere ayrılarak yatırımcılar için risk minimizasyonu ve stratejik yatırım gibi farklı açılardan çeşitli senaryolar geliştirilmektedir. Hisse hareketlerinin analizinde SAX kümeleme yöntemi kullanılmaktadır. Ayrıca dendrogram ağaç grafiği, farklı SAX kombinasyonlarının kümelenmesini analiz etmek için kullanılır.

Anahtar kelimeler: Makine Öğrenmesi, Sembolik Toplam Yaklaşımı (SAX), BIST100, Borsa

Acknowledgements

I would like to express my sincere gratitude to my advisors. Assoc. Prof. Dr. Ramazan Ünlü and Dr. İsmet Söylemez for his support and guidance in all my research that started from my B.Sc. degree and has continued until now. Also, I would like to thank jury members Asist. Prof. Dr. Ahmet Erdoğan and Asist. Prof. Dr. Betül Çoban for their contributions to my thesis.

I'm extremely grateful to my parents Ayhan Nalici and Esma Nalici, this endeavor would not have been possible without them, and especially to my brother Levent Emre Nalici for constantly reading the parts finished and giving feedback to idealize the written expressions of the study.

Additionally, I would like to thank my friends İbrahim Tümay Gülbahar and Sami Kaya for their supports.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

To My Family

Chapter 1

Introduction

The goal of financial management is to maximize shareholder value. Maximizing the value of the company also means maximising the wealth of those who invest in the company. This fundamental objective can be achieved by maximising the market value of the company's existing shares. To achieve this goal, companies use market capitalisation, profitability, earnings per share, price/earnings ratios, etc. They aim to maximize the market value of the company by closely monitoring the prices. Market value of a company traded on BIST it is the value obtained by multiplying the closing price of the shares on the stock exchange for the day analysed by the total number of shares of the company [1]. The stock market has rapidly developed in our country in recent years and serves as a financial resource for companies. It is considered an alternative investment tool for investors who want to invest their savings. Rational investors always aim to earn higher returns, and correct determination of security prices is crucial in calculating returns. The information factor is one of the most fundamental factors that affect security prices. In active markets, new data or information is analysed and evaluated by market actors, forming a new market price for the security in question. This market equilibrium price persists until new information is introduced [2].

One of the key goals of market stakeholders like traders, investors, and market makers is forecasting stock market returns. They will develop buy-or-sell strategies based on their projections and will employ technical and basic research to make forecasts. Latest study has demonstrated that, from the standpoint of an investor, the sign predictability of stock price returns is both feasible and profitable. Because of the market's unpredictability and inherent risk, it is imperative that innovative technology be used to produce workable solutions [3]. Stock prices are influenced by a wide range of factors such as including politics, investor sentiment, supply and demand, and natural disasters. Computational finance has focused on exploiting massive financial data, especially real-time information, to anticipate stock price movements using machine learning and neural

network methods, in addition to financial economics and time series monetary economics [4].

Several applications in computing are built on data analysis, either during their development phase or as part of their online operations. Depending on the availability of suitable models for the data origin, methods for analysing data might be classified as descriptive or confirmatory; however, grouping or classifying measurements based on either goodness-of-fit to a proposed model or clustering discovered through analysis is a crucial step in either hypothesis formation or decision-making. A set of patterns are grouped together using cluster analysis based on their degree of similarity [5]. Clustering is an unsupervised machine learning technique. The process of extracting examples from datasets of input data without labelled replies is known as unsupervised learning. To extract useful information from unlabelled data, clustering has mostly been utilized as a statistical technique to group such data. The objective of clustering is to divide the population or set of data points into several groups so that the data points within each group are more like one another and different from the data points within the other groups. It is essentially a grouping of items based on how similar and unlike they are to one another. The definition of clustering is the grouping of things in which there is little to no information of the links between the items in the information that is given. Clustering requires to make the root classes in the data visible. Furthermore, clustering is a technique that divides unlabelled data into distinct classes with a minimum of oversight [6]. Hard clustering and soft clustering are the two main categories of clustering algorithms. A data point can belong to two or more groups when using soft clustering as opposed to hard clustering, where a data point can only belong to one cluster. Hierarchical algorithms and partitional algorithms are two categories of hard clustering methods. In the case of a partitional algorithm, the dataset is divided into a single partition. In contrast, a sequence of divisions is created inside the dataset when using hierarchical techniques [7]. Moreover, since there isn't a label associated with the patterns in clustering, it is thought to be more challenging than supervised classification. In the case of supervised categorization, the assigned label serves as a hint for categorizing data items. On the other hand, clustering makes it challenging to determine which group a pattern will belong to in the absence of a label. Numerous variables or traits may be deemed appropriate for clustering. The situation could be made worse by the curse of dimensionality. large computational cost and large dimensionality both have an impact on the consistency of

algorithms. However, feature selection techniques have been suggested as a solution. The clustering criteria may also be influenced by the database sizes [8].

The appropriate model choice has a significant impact on the simplicity and effectiveness of time series data mining, as it does with most computer science issues. The available algorithms, data structures, and terminologies are constrained as a result. Due of these restrictions, academics are thinking about representing time series symbolically [9]. A time series of any length can be converted by Symbolic Aggregate Approximation (SAX) into a string of any length. The size of the alphabet is also an undefined number. Because it utilizes an illustration in between the raw time series and the symbolic strings, the discretization process is distinctive. The dataset first approximates the data using the Piecewise Aggregate Approximation (PAA) representation, which is then symbolized as a discrete string as a letter [10].

In this study, the data of the "BIST 100" companies between 01.01.2020-08.08.2023 years are used. This study clusters BIST100 companies using the Symbolic Aggregate Approximation (SAX) method and provides investors with various scenarios for risk minimization and strategic investment. The SAX method helps investors make more informed and sound decisions by autonomously identifying companies that exhibit similar price movements. The contribution of this model to the literature is its effective use in uncovering hidden patterns from unlabelled data, thus providing important strategic guidance in investment decisions.

The remainder of the thesis structured as follows. In “Chapter 2”, a literature review regarding the Symbolic Aggregate Approximation (SAX) method and clustering techniques. In addition, the explanation of the "SAX" method and the formulas used are included. In “Chapter 3”, details of the data set used in the thesis, pre-processing methods and how the "SAX" method is applied are explained. Moreover, the results of the model which is explained in “Chapter 3” described and the analysis of these results are explained. Chapter 4 provides a summary and discussion of the societal impact and contribution to global sustainability of the thesis. The chapter concludes by outlining further directions for related research.

Chapter 2

Literature Review

2.1 Clustering

The goal of clustering is to make the primary groups in the data more visible. Additionally, clustering is an algorithm that divides unlabelled data into distinct classes with a minimal guidance. The arrangement of the objects ensures that objects belonging to the same class share traits and differ from those belonging to other classes. Another way to explain clustering is as a machine learning component that deals with unsupervised learning. Algorithms that detect characteristics from datasets derived from either real or simulated data represent the learning process [6]. Algorithms for clustering are often categorized as either using a hierarchical or partitioning technique to get results.

2.1.1 Hierarchical Clustering

Clusters are created using top-down or bottom-up approaches to repeatedly divide the patterns in hierarchical clustering algorithms [11]. Agglomerative and divisive hierarchical clustering are the two types of hierarchical methods. The bottom-up approach used by the agglomerative build's clusters from the ground up, starting with atomic clusters of a single object and merging them into larger and larger clusters until all the objects are eventually contained within a single cluster or until other termination conditions are met. The divisive hierarchical clustering uses a top-down method to split up large clusters of items into smaller clusters, which it continues to do until each object becomes its own cluster or until it meets termination criteria. The single-link, complete-link, and minimum-variance algorithms are the most common incarnations of hierarchical clustering algorithms. The most widely used of these are the single-link and complete link algorithms. The way these two techniques describe the similarity between two clusters is different. The smallest distance between any two patterns produced from the two clusters is the distance between two clusters in the single-link approach. The distance that exists

between two clusters in the complete-link method is equal to the sum of all pairwise distances between patterns in the two clusters. Based on minimal distance requirements, two clusters are combined to create a bigger cluster in either scenario. Clusters that are compact or closely bonded are produced using the complete-link algorithm. On the other hand, the single-link algorithm experiences a chaining effect [5].

2.1.2 Agglomerative Hierarchical Clustering

In terms of algorithms, agglomerative hierarchical clustering methods can be described as greedy. To create the appropriate data structure, an irreversible algorithm is utilized in a series of stages. Assume that each stage of the process involves merging or aggregating a pair of clusters, maybe containing singletons. Numerous agglomerative hierarchical clustering techniques have been put forth at various points in time [11]. It is reasonable to divide these hierarchical algorithms into two categories of techniques. The single, full, weighted, and unweighted average linkage methods make up the first category of linkage techniques. These are techniques that can benefit from a graph representation. The second category of hierarchical clustering techniques includes those that let you specify the cluster centres. The centroid, median, and minimal variance procedures are some of these techniques [12].

2.1.3 Divisive Hierarchical Clustering

The Divisive Hierarchical Clustering usually starts with all the objects in the same cluster. The use of K-means Clustering then divides a cluster into individual clusters throughout each successive iteration. Until every object in a cluster is down or the termination condition is reached, the system is offline. This approach is rigorous; once merging or splitting has been carried out, it cannot be reversed [13]. All items are first grouped together into a single sizable cluster in any divisive hierarchical clustering process. A cluster is further split into two with each cycle. The principle governing how to separate or divide the cluster is what is important [14].

2.1.4 Partitional Clustering

The iterative relocation technique, commonly known as partitional clustering, is thought to belong to the most common class of clustering algorithms. These algorithms iteratively move data points across clusters until an ideal partition is reached to minimize a specific

clustering criterion. The data points are divided into k partitions by the partition clustering technique, with each partition denoting a cluster. A particular goal function is used to partition the data. The items inside a cluster are similar, but the objects of separate clusters are "dissimilar", and the clusters are generated to maximize an objective partitioning criterion, such as a dissimilarity function based on distance. Applications requiring a fixed number of clusters can benefit from partitioning clustering techniques [13].

2.1.5 K-Means Clustering

One of the most well-known, widely used, and simplest clustering methods is the k-means technique, which is frequently used to address clustering issues. The provided data set is categorized in this technique using a user-defined number of clusters, k. To define k centroids, one for each cluster, is the main notion [14]. Until there are no more items that can be assigned to clusters, the algorithm continuously changing the assignment of objects to the nearest current cluster mean. The simplicity of this method is one of its benefits. It also has a few shortcomings. It is quite challenging to predict the number of clusters in advance. It is susceptible to outliers since it deals with squared distances. The centroids' lack of use in solving most issues is another negative [13].

2.1.6 Partitioning Around Medoids (PAM)

The clustering technique Partitioning Around Medoids (PAM) uses the k-medoids model. Like k-means clustering, partitioning around Medoids is referred to as partitional clustering. PAM clustering employs data points, which have a smaller total distance of the resulting grouping, as opposed to k-means clustering, which uses the mean of the data points inside cluster to become cluster centre [15]. Moreover, the technique uses the same stages as the k-means algorithm, but instead of using means, it uses medoids, which makes it more resistant to outliers. PAM may be used to datasets that contain category data as well as other discrete data types, such binary data. The need that the intended number of clusters be preset is one issue with the PAM method [13].

2.1.7 Clustering Large Applications (CLARA)

Groups of clusters with comparable geometric features are created using the Clustering Large Applications (CLARA) clustering method. CLARA is intended to group at least 100 offers different algorithms for less things [16]. Furthermore, CLARA is introduced to address the PAM problems. Unlike PAM, this operates on a larger data collection.

Instead of using the entire dataset, this approach merely uses a sample of the data. Using the PAM method, it determines the medoid along with the information at arbitrary [17].

2.1.8 Clustering Large Applications Based on Randomized Search (CLARANS)

PAM and CLARA are comparable to CLARANS. The selection of medoids is done at random to start. It interactively sketches the neighbour. “Max Neighbour” is checked for exchanging. A different medoid set is used if the pair is negative. If not, it selects the current selection of medoids as the local optimum and then randomly selects a new set of medoids. The procedure is stopped till the best is returned. CLARA is introduced to address the PAM issue. Unlike PAM, this operates on a bigger data collection. Instead of using the entire data set, this approach merely uses a sample of the data. Applying the PAM method, it determines the medoid along with the information at arbitrary [17].

2.2 Symbolic Aggregate Approximation (SAX)

The Symbolic Aggregate Approximation (SAX) method is an approximation method used to reduce the amount of data used in time series. Since it utilizes a representation between the raw time series and the symbolic characters, this discretization process is distinctive. The data is first translated into a discrete string representation called a Piecewise Aggregate Approximation (PAA) model. These features are representative data are then converted into symbols. The algorithm has two important features. These are "Dimensionality Reduction" and "Lower Bounding" [9, 10].

2.2.1 Dimensionality Reduction

The collection of time series that make up the database are designated as $Y = \{Y_1, Y_2, \dots, Y_k\}$, and a time series query is indicated as $X = \{X_1, X_2, \dots, X_n\}$, The length of each sequence in Y is n units without losing generality. Let N represent the converted space's dimensions ($1 \leq N \leq n$) so that can be index it. To make things easier, N is a factor of n. The method is not needed this, although it does make the notation simpler [18]. Equation (2.1) is the mathematical representation of dimensional reduction.

$$\bar{X}_{\mathrm{i}} = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\left(\frac{n}{N}\right)_i} X_j \qquad \textbf{(2.1)}$$

Moreover, a dataset is represented as an n-dimensional ($n\ x$ D) matrix X_i, made up of n data vectors (i ∈ {1, 2, …, n}). Let's further assume that this dataset has inherent dimension d, where d is often greater than or equal to D. The points in dataset X are located on or close to a surface with dimensionality d that is situated in the D-dimensional space, demonstrating whatever is meant by inherent dimensionality in this context. Dataset X with dimensions D is converted into a new dataset Y with dimensionality d using dimensionality reduction techniques, preserving as much of the original data's shape as feasible. In broadly, neither the inherent dimensionality d of the dataset X nor the shape of the data surface can be determined [19]. For the Dimensional Reduction Piecewise Aggregate Approximation technique is used before symbolizing dataset.

2.2.2 Piecewise Aggregate Approximation (PAA)

The Piecewise Aggregate Approximation (PAA) methodology is a standard and quick real-value technique that represents an ordered series using mean values of equal-sized sections [18, 20]. Furthermore, it alternately depicts a time series by dividing it into intervals and substituting the mean value for each one. The technique's goal is to reduce the number of points and clutter in a time series while maintaining the trend [21]. It is generally known that comparing time series with various delays and intensities is pointless, thus before transferring each time series to the PAA form, it is normalized to have a mean of zero and a standard deviation of one [10]. In Figure 2.1, an example of PAA application is shown. Blue dots indicate normalized time series data. The yellow line indicates the data set after PAA is applied. In this example, our original data consists of 60 points. If we take the number of PAA segments as 6, a single value is determined by averaging 10 sequential data. In short, this averaged data represents 10 sequential data. Thus, the data set will be reduced from 60 data to 6 data and the characteristics of the data will be preserved.

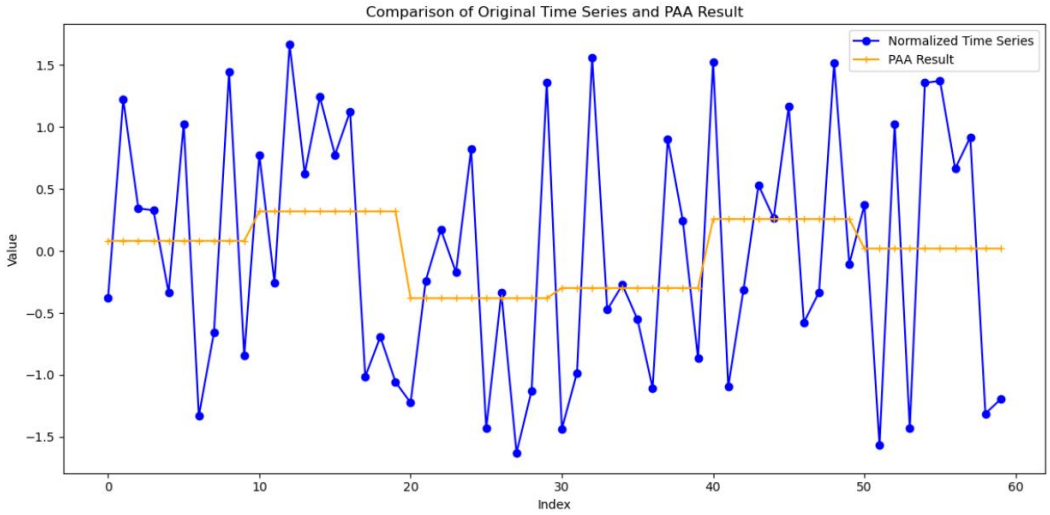


Figure 2.1 Example of PAA application for n=60 and Number of Segment = 6

2.2.3 Lower Bounding

It takes some work to demonstrate that a distance measure between two symbolic strings lowers the real difference among the original time series. The crucial finding that enabled us to establish lower limits was that the PAA distance measurement isn't as limited by the symbolic distance measure. Then, by only citing the previous justifications for the PAA representation itself, we may use transitivity to demonstrate the intended outcome. Additionally, there is a compromise among the variable a, which controls the level of detail of each estimating item, and the value for "number of segments," which controls the number of approximating elements. The optimum compromise cannot be determined mathematically since it depends heavily on facts. However, using a straightforward experiment, we can empirically identify the optimal values. Since we want to obtain the tightest lower limits possible, we can simply estimate the lower bounds over all viable parameter ranges and select the most suitable values [10].

2.2.4 Discretization

After converting a time series database into the PAA, one more conversion might be used to provide a discrete representation. A discretization method that generates symbols with equiprobability seems required. Since normalized time series have a Gaussian distribution, this is simple to accomplish. The SAX requires that all the track's values are distributed normally. Based on this assumption, it computes quantiles to quantify the values and classify them into symbols, distributing the data points across the symbols

somewhat evenly. This makes it easier to focus on the values' concentrated value ranges [22].

2.3 Distance Measure

Different distance measurement techniques can be used to measure the difference between two data. If we minimize this amount of distance, it can be concluded that the two data are the same. On the other hand, we can assume that the distance between the two datasets is the same.

2.3.1 Euclidean Distance

The "the-crow-flies length" is a measurement used in Euclidean distance. This formula is used to calculate the length between two points, $X \in \{X_1, X_2 \ldots\}$ and $Y \in \{Y_1, Y_2 \ldots\}$. Calculating the square root of the sum of the squares of the discrepancies between identical quantities is necessary to get the Euclidean distance across each of the data points [23] . Equation (2.2) represents the Formulation of Euclidean Distance.

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \quad \textbf{(2.2)}$$

2.3.2 Manhattan Distance

The Manhattan distance calculation determines the length of the route that needs to be taken along a grid to go from one data point to the next. The total of the gaps between two objects' related components is known as the Manhattan distance. This distance is calculated using the formula between the points X ∈ (X1, X2, …) and Y ∈ (Y1, Y2, …) [23]. Equation (2.3) represents the Formulation of Manhattan Distance.

$$Manhattan\ Distance = \sum_{i=1}^{n} |X_i - Y_i| \quad \textbf{(2.3)}$$

2.3.3 MINDIST Distance

The "MINDIST" distance is used for the symbolized datasets after the "PAA" transformation. This function's objective is to calculate the separation between two SAX representations of a time series, which necessitates calculating the separations among every set of symbols. This formula uses the value corresponding to each letter in the normal distribution to measure distance (“dist” function). These values come from the normal distribution divided into equal areas by alphabet size. A separate function is defined for these values [10] .Then, the difference of each letter in the two "SAX" series to the letter in the other series was calculated. These values are squared. Then, this process was performed for the values in the whole series, and these values are summed, and their square roots are taken. Lastly, this value is multiplied by "Compression Value". The "Compression Value" is obtained by dividing the length of the dataset by the length of the "PAA" transformation of the dataset and taking its square root. Equation (2.4) shows the calculation of Compression Rate and Equation (2.5) represents the calculation of MINDIST distance.

$$Compression\ Rate = \sqrt{\frac{Length\ of\ Time\ Series}{Length\ of\ PAA\ Time\ Series}} \tag{2.4}$$

$$MINDIST = Compression\ Rate\ x \sqrt{\sum_{i=1}^{Length\ of\ PAA\ Series} (dist)^2} \tag{2.5}$$

2.4 Symbolic Aggregate Approximation Application

There are different studies in the literature using SAX clustering applications. As examples of these studies, SAX has been used to show that the location identification approach based on the product pipeline transportation system is feasible [24], choosing an accurate time frame for power system planning [25], using simulation models for garment production lines [26], time series clustering in industrial application systems [27], dimensionality reduction in time series in Industrial 4.0 application [28], analysing

of Financial Time Series [29, 30] ,Stock Portfolio Optimization [31]. Moreover, there are different types of “SAX” in the literature such as SAXRegEx [22], SAX-ARM [32, 33], Extreme-SAX [34], HOT SAX [35-38], Extended SAX [39], SAX Parameter Estimation [40].

Table 2.1 shows the data sets to which 'SAX' was applied in the selected articles from the most cited and recently published articles according to the Web of Science database. The purpose of the data sets for which the 'SAX' process is applied is also indicated. For instance, the "SAX" method is used in classification methods. In 10 of the studies listed in Table 2.1, "UCR Datasets" data is used when preparing data for the classification method. [41-50] studied for classification performance. They worked on improving the classification performance of different data in the UCR data set with different methods they developed. In other studies, classification studies are carried out with different data sets such as energy [51, 52], health [53-56] , finance [57, 58], weather [59-61], traffic [61], bearing [62, 63], network [64] and manufacturing [65-71]. Moreover, the number of articles on clustering is very small compared to the number of articles on classification. There are 7 articles about clustering. 4 of these articles are about clustering energy and electricity [72-74] consumption data and they tried to identify consumers with the same consumption habits using different methodologies. The other two articles aimed to bring together people with the same habits, but they worked on different data sets [75-77]. Furthermore, there are 9 articles that use the "SAX" model on prediction. While 2 of these articles worked on the prediction of energy consumption, 1 worked on stock market data, 1 worked on maintenance data and the other worked on traffic data, 2 worked on prognostic dataset and the other ones worked on traffic data and COVID datasets.[78-86]. Other studies in Table 2.1 used the "SAX" method to perform studies such as "Boundary Detection", "Anomaly Detection", "Motif Discovery" [87-100].

Table 2.1 Summary of Literature

DATA TYPE	PURPOSE	PAPER
UCR Datasets	Classification	[41]
UCR Datasets	Classification	[42]
UCR Datasets	Classification	[43]
UCR Datasets	Classification	[44]
UCR Datasets	Classification	[45]
UCR Datasets	Classification	[46]
UCR Datasets	Classification	[47]
UCR Datasets	Classification	[48]
UCR Datasets	Classification	[49]
UCR Datasets	Classification	[50]
Power Consumption Data	Classification	[51]
Power Meter Sensors Data	Classification	[52]
Intensive Care Unit (ICU) & Hepatitis Dataset	Classification	[53]
Electroencephalography (EEG) Data	Classification	[54]
Devices, ECG Signals Human Motion Simulation and Spectrum-measurement datasets	Classification	[55]
Electrodermal Activity Signals Dataset	Classification	[56]
Stock Price Dataset	Classification	[57]
World Bank Datasets	Classification	[58]
Vehicle Trajectory & Weather Data	Classification	[59]
Traffic Detector Data & Weather Data	Classification	[60]
IEEE 24 Bus System Data & IEEE 118 Bus System Data	Classification	[61]
Bearing Data	Classification	[62]
Bearing Data	Classification	[63]
Network Security Datasets	Classification	[64]
Soil Moisture Datasets	Classification	[65]
Manufacturing Data	Classification	[66]
Landsat Time Series Data	Classification	[67]
Calibration Dataset	Classification	[68]
Drilling Industry Data	Classification	[70]
Vibration Dataset	Classification	[71]
Electricity Consumption Data	Clustering	[72]
Electricity Consumption Data	Clustering	[73]
Energy Consumption Data	Clustering	[74]
Twitter Stream Data	Clustering	[75]
Human Mobility Data	Clustering	[76]

Rail Passenger Dataset	Clustering	[77]
Stock Price Dataset	Prediction	[78]
Energy Consumption Data	Prediction	[79]
Energy Consumption Data	Prediction	[80]
Pipe Skid Maintenance Data	Prediction	[81]
Performance Measurement System Dataset & Taxi Tracks Traffic Speeds Dataset	Prediction	[82]
COVID Disease Datasets	Prediction	[83]
Prognostic Datasets	Prediction	[84]
NASA Ames Prognostics Data	Prediction	[85]
Environmental, Financial, Industrial, Health Datasets	Prediction	[86]
Energy Performance Dataset	DayFilter Process	[87]
Crop Monitoring Data	Extracting Information	[88]
Building Automation System Data	Anomaly Detection	[89]
Electricity Consumption Data	Anomaly Detection	[90]
Power-Quality Waveform Dataset	Boundary Detection	[91]
Traffic Data	Quality Analysis	[92]
Entomologists Data	Motif Discovery	[93]
Electricity Consumption Data	Anomaly Detection	[94]
Power Consumption Data	Anomaly Detection	[95]
Electric Vehicle Driving Data	Anomaly Detection	[96]
Resilient Distributed Datasets	Motif Discovery	[97]
Keyboard and Mouse Interactions Dataset	Anomaly Detection	[98]
S&P500 Data and Open Power System Data	Motif Discovery	[99]
Asteroids' orbital elements Data	Motif Discovery	[100]
Electricity Consumption Data	Clustering	[101]
Threshing Cylinder Data	Classification	[102]

SAX method makes analyses more efficient by reducing data size and filtering noise in financial data sets. By summarising time series in symbolic form, it facilitates the detection of complex patterns and anomalies. This reduces computation time and optimises resource utilisation, especially for large volumes of financial data. In addition, with SAX, data mining and machine learning algorithms can run faster and more effectively. To better model and predict fluctuations and trends in financial markets, the SAX method provides advantages by preserving the data structure and minimizing information loss. SAX method makes analyses more efficient by reducing data size and filtering noise in financial data sets. By summarising time series in symbolic form, it facilitates the detection of complex patterns and anomalies. This reduces computation

time and optimises resource utilisation, especially for large volumes of financial data. In addition, with SAX, data mining and machine learning algorithms can run faster and more effectively. To better model and predict fluctuations and trends in financial markets, the SAX method provides advantages by preserving the data structure and minimising information loss.

Chapter 3

Material and Methods

3.1 Data

BIST100 companies' data are used for analysis. In this research, closing data of companies in BIST100 between 2nd January 2020 and 8th August 2023 are used. Table 3.1 shows the examples of close stock prices of companies. When the data is examined before starting the analysis, it is determined that AHGAZ, ALFAS, ASTOR, AYDEM, BIOEN, CANTE, EUREN, GENIL, GESAN, GWIND, KCAER, KMPUR, KONTR, KZBGY, PENTA, PSGYO, QUAGR, SMRTG, YYLGD companies have missing data between these dates. To maintain data accuracy, 81 companies, excluding those with missing data are used for analysis. As a result, a dataset consisting of 81 columns and 900 rows is used for SAX analysis.

Table 3.1 Example Raw Dataset of BIST100 Close Price

Date	AEFES	AGHOL	AHGAZ	…	VESTL	YKBNK	YYLGD	ZOREN
2020-01-02	16.39649	16.79828	NA		10.1583	2.19668	NA	1.65
2020-01-03	16.29712	16.27363	NA		10.44178	2.135662	NA	1.59
2020-01-06	15.94222	15.7776	NA		10.39454	2.092077	NA	1.52
2020-01-07	16.65202	16.16871	NA		10.81977	2.092077	NA	1.56
⋮					⋮			
2023-08-03	100.400002	144.000000	12.75		108.699997	4.950000	86.800003	38.279999
2023-08-04	102.099998	152.199997	12.51		118.599998	5.120000	88.699997	38.259998
2023-08-07	107.699997	167.399994	12.60		124.599998	5.170000	90.300003	39.259998
2023-08-08	105.000000	163.500000	12.62		116.900002	4.900000	87.599998	38.340000

When the distribution of the remaining companies is examined, it is observed that there are companies from 9 different sectors. The main sectors of the companies reported on Public Disclosure Platform are used [103]. According to Table 3.2, distribution of the 81 companies, 3 are "Mining and Quarrying", 32 are "Manufacturing", 4 are “Electricity, Gas and Water”, 1 is "Construction and Public Works", 7 are "Wholesale and Retail Trade", 2 are "Transportation and Storage", 25 of them provide services in the "Financial Institutions", 1 in the "Technology", and 2 in the "Information and Communication" sector.

Table 3.2 Sector Based Company Numbers (81 Company)

SECTORS	NUMBER
Mining and Quarrying	3
Manufacturing	32
Electricity, Gas and Water	4
"Construction and Public Works	1
Wholesale and Retail Trade	7
Transportation and Storage	2
Financial Institutions	25
Technology	1
Information and Communication	2

3.2 Standardization

Data normalization is one of the methods of preliminary processing in which the data is scaled or changed to ensure that each characteristic contributes equally. The quality of the data needed to create a generalized prediction model for the classification issue determines how well machine learning algorithms perform. Numerous research has demonstrated the significance of data normalization for enhancing data quality and consequently the effectiveness of machine learning algorithms [104]. Moreover, cleaning of information, cooperation, modification, and reduction are all parts of data preparation, which has as its major objective ensuring the quality of the data before it is supplied to any learning algorithm. The z-score, decimal scaling, and min-max normalization techniques are the most often used standardization techniques for transforming data [105]. In this study, z-score technique is used for normalization. Data is scaled for standard normal distribution which is the mean is zero and standard deviation is 1. For this purpose, the mean and standard deviation of each company are calculated separately. After the data reduction process of each company is made, it is brought together again for SAX analysis as shows in Table 3.3. Normalized calculation of Stock Price for each company is

calculated via Equation (3.1). X denotes stock price. μ is mean stock price of company and σ is standard deviation of stock price of company.

$$\frac{X-\mu}{\sigma} = X_{normalized} \tag{3.1}$$

Table 3.3 Example Normalized Dataset of BIST100 Close Price For 81 Companies.

Date	AEFES	AGHOL		VESTL	YKBNK	ZOREN
2020-01-02	-0.78889	-0.85898		-1.13744	-0.6906	-0.7406
2020-01-03	-0.79415	-0.87458	…	-1.1198	-0.70832	-0.77979
2020-01-06	-0.81296	-0.88932		-1.12274	-0.72097	-0.82551
2020-01-07	-0.77535	-0.8777		-1.09629	-0.72097	-0.79938
⋮						
2023-08-03	3.661844	2.922657		1.91021	2.754363	1.003234
2023-08-04	3.751915	3.166439		1.938203	2.870518	0.98364
2023-08-07	4.048618	3.618326	…	2.103053	2.980865	0.944453
2023-08-08	3.905565	3.502381		1.981749	3.007	0.898735

3.3 Piecewise Aggregate Approximation (PAA)

Time series data mining has attracted considerably more attention with the growth of temporal data sets from many fields, for dimensionality reduction and the creation of effective similarity metrics, representations are crucial. High-level representations such as Fourier transforms, wavelets, piecewise polynomial models, are taken into consideration. To reflect the similarity of the time series, autoregressive kernels are recently implemented [106]. A major challenge in the processing and administration of time series data is similarity search and detection. While several dimensionality reduction strategies have been proposed to increase the effectiveness of similarity searches, most approaches to this problem have been developed around the idea of dynamic temporal warping. There is a critical need for enabling similarity identification in time series in a method that is both accurate and quick due to the constant growth of sources of time series data and the importance of real-world applications that use such data [107].

PAA method is used for dimensionality reduction. Applying the mean values of sections of similar size, the PAA methodology expresses an ordered series quickly and

conventionally [108]. Additionally, alternatively represent a time series by splitting it into intervals and replacing each one with the mean value. To retain the trend, the approach aims to decrease the number of points and clutter in a time series [21].

In Figure 3.1, it is an example of application of PAA on the data. Blue line represents the normalized values of AEFES company closed stock price. Orange line represents the PAA results of AEFES. In this example, dimension is reduced to 900 to 10. Thus, the average of the 90-day closing data is taken and the average of the 90-day data is assumed as a single data.

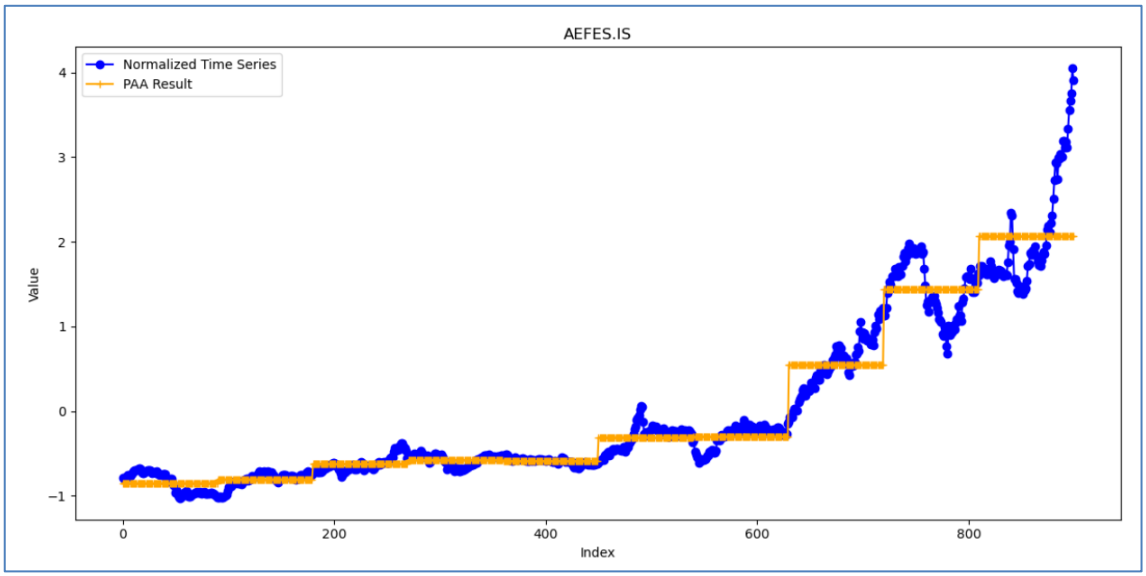


Figure 3.1 The Example of Piecewise Aggregate Approximation of AEFES For Number Segment=10

3.4 Symbolic Aggregate Approximation (SAX)

In time series, the SAX approach is a approximation technique. This discretization approach differs because it employs a representation intermediate between the raw time series and the symbolic characters. A symbolic transformation that turns a time series into a string is built into SAX on top of PAA. The time series is segmented, just as PAA, except the lengths of the segments might vary. To find the alphabet symbol that each segment is mapped to, a table of breakpoints and the mean of each segment are combined [109]. PAA model, a discrete string representation of the data, is created first. Then, the PAA applied dataset is assigned to the specified alphabet number. For this, first the alphabet number is determined. The values that will allow the normal distribution graph to be divided into equal areas are determined as critical values. Letters are assigned to

these fields in alphabetical order. Whichever range the PAA value of data corresponds to, the letter assigned to this range will be the letter used in place of the value [9, 10].

In Figure 3.2, an example of the application of SAX on the data is shown. The PAA version of the "AEFES" data in Figure 3.1 is used. In this example, the Normal distribution graph is divided into 6 equal parts which is alphabet size. As a result, the critical points are -0.97, -0.43, 0, 0.43 and 0.97, respectively. Finally, alphabet letter assignment is made according to which range the data corresponds to, and the vector [b, b, b, b, b, c, c, e, f, f] will be used instead of the actual data.

Figure 3.2 The Example of Piecewise Aggregate Approximation of AEFES For Number Segment=10 and Alphabet Size=6

3.5 Clustering

Making the major groupings in the data more obvious is the aim of clustering. In addition, clustering is an algorithm that, with the least amount of direction, separates unlabelled data into discrete groupings. The placement of the items ensures that they have characteristics in common with one another and set them apart from objects of other classes. Clustering may also be thought of as a kind of machine learning that deals with unsupervised learning. The learning process is represented by algorithms that find features in datasets created from either actual or simulated data [6]. To achieve outcomes, clustering algorithms are frequently characterized as either employing a hierarchical or partitioning method.

In this study, dendrogram tree graph, which is one of the hierarchical clustering methods, is used. This involves employing certain algorithms designed to create a dendrogram which is a form of tree diagram that displays how clusters created by a clustering technique are arranged. These methods use the given input data and generate a dendrogram-able hierarchical structure of clusters. Cluster analysis is a type of hierarchical grouping. The hierarchical structure of hierarchical algorithms, which is established throughout the algorithm's execution, is one of the primary benefits over non-hierarchical alternatives. Every hierarchy in the dendrogram represents a single algorithm step [110].

The letter vectors of the companies created using the SAX algorithm were used as the input of the dendrogram graph. The "Unicode" values of the letters are used to measure the distance between the letters. Thus, it is aimed that the same letter vectors are in the same cluster. The "Euclidean" distance is used when measuring the distance between vectors. Moreover, different linkage method is used for creating clusters such as “Single”, “Average”, “Complete” and “Ward”. The lowest distance between a particular instance from the first cluster and an instance from the second cluster is how one linkage determines the distance between two clusters. While the measurement of the farthest neighbour contributes to the impact of marginal data, complete linkage marginal situations prevent near clusters from merging. To offer a more realistic assessment of the distance between clusters, average linkage is intended to provide a natural compromise between the linkage measurements. The distances between each instance in the first cluster and each case in the second cluster are computed, averaged, and then used to determine average linkage [111]. Every conceivable union of clusters is considered at each stage of the Ward clustering process, and the two clusters that result in the least amount of information loss when they merge are joined. The Sum of Square criteria was used by Ward to characterize information loss [112]. The objective is to locate the two clusters that are closest to one another and unite them. There are several distinct connection metrics that each define the distance between cluster pairs differently. Some metrics use the minimum or the largest range that can be discovered between pairs of examples where each of them is from a separate cluster to determine the distance between two clusters [111]. Finding the ideal number of clusters is an important process in clustering algorithms. Different methods can be used to find the ideal number of clusters. In this study, the "Elbow Method" is used. The elbow technique examines the amount of variation that can be explained as a function of the number of clusters. This strategy

assumes that one should select several clusters to ensure that adding more clusters does not significantly improve the modelling of the data. Plotted against the number of clusters is the proportion of variation explained by the clusters. The first clusters will significantly increase the amount of information, but at some point, the marginal gain will drastically decrease, giving the graph an angle. The "elbow criterion" refers to the decision to select the appropriate number of clusters at this stage. The cost begins to decline sharply at a certain number for k, and when you raise it more, it then comes to a rest [113]. "Distortion" value is used as "Key Performance Index" in the elbow method. Small k values are ideal for the elbow approach. The elbow technique figures out the squared variation between several k values. The average distortion degree decreases as the k value rises. Each category has fewer samples, and the samples are located nearer the centre of gravity. The k value corresponding to the elbow is the place where the improving impact of the distortion degree drops the fastest as the k value rises [114].

3.6 Experimental Study

Normalized data of 81 companies are used for analysis. For the application of the SAX algorithm, tests are carried out on different alphabet sizes and segment numbers. Each of the alphabet numbers from 3 to 10 is tried with different segment numbers. Moreover, segment numbers are chosen as values that give a mod 0 result of 900 which is the size of the data. Since it is difficult to analyse all combinations for validation, the following combinations are chosen as sample: (Alphabet Size “4” and Number of Segment “25”), (Alphabet Size “9” and Number of Segment “25”), (Alphabet Size “7” and Number of Segment “45”), (Alphabet Size “7” and Number of Segment “180”). SAX application is performed for each combination and a dendrogram is plotted using the vectors converted to "Unicode".

Figure 3.3 represents the example of dendrogram for Alphabet Size “4” and Number of Segment “25” with usage of Ward linkage. The names of the companies are shown on the x axis. Furthermore, on the y axis, it shows the distances between vectors. Horizontal lines indicate companies that form clusters together according to the distance specified on the y-axis. Each vertical line intersecting each horizontal line represents different groups. For instance, in Figure 3.3 represents the 4th cluster for Alphabet Size “4” and

Number of Segment “25” with usage of Ward linkage. If we take the dashed line as reference, 4 clusters can be formed with approximately 7.2.

A comparison of different cluster numbers is made for each alphabet size, number of Segment combination. The purpose of this analysis is to identify companies that act together in different cluster numbers. The results of 10, 20, 30 and 40 clustering are analysed using the dendrogram graphics of the combinations determined because of the tests. Moreover, Cluster values selected using the elbow method are also analysed in the dendrograms. The elbow method is analysed for each alphabet and segment number combination and different linkage methods. In figure 3.4, represents the example application of Elbow Method for Alphabet Size “4” and Number of Segment “25” with usage of Ward linkage. In this figure, x axis represents number of clusters from 1 to 50 and y axis represents the amount of distortion for number of clusters. When the graph is analysed, the ideal cluster value is determined as 13 and the "distortion" value is 161.965. Dendrograms of other linkage methods and other SAX parameters are listed in the Appendix A.

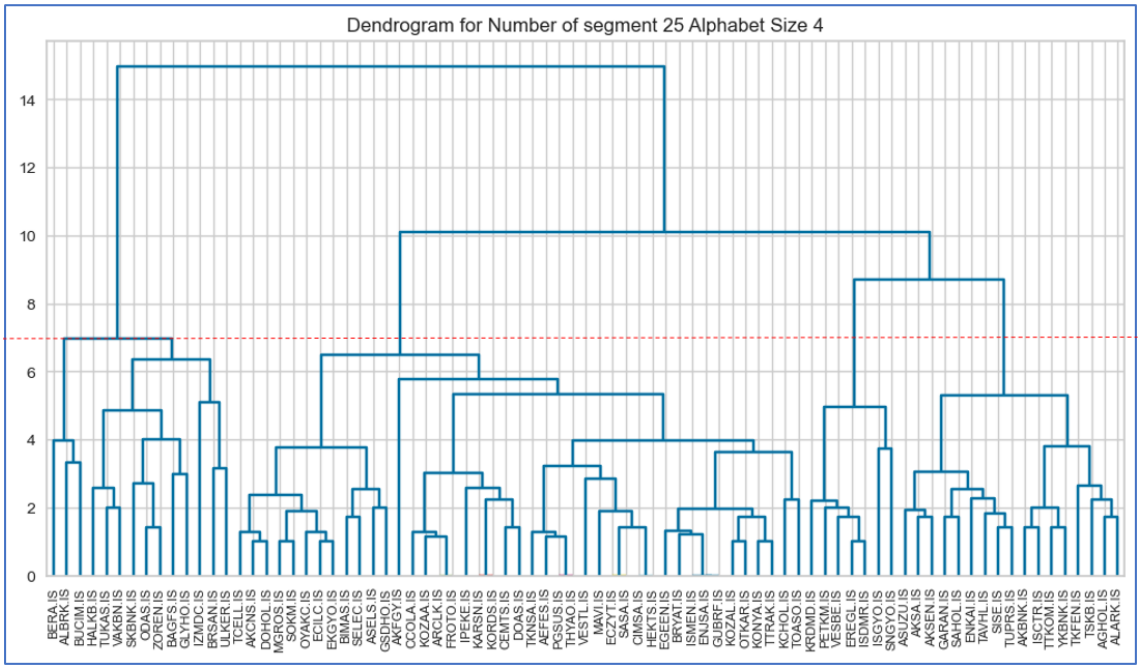


Figure 3.3 The Example of Dendrogram for Alphabet Size = 4 Number of Segment = 25 With Usage of Ward Linkage

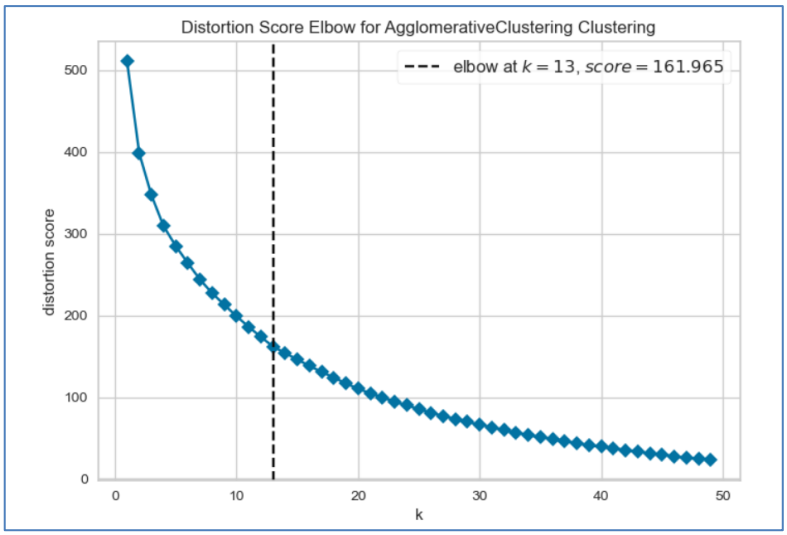


Figure 3.4 The example application of Elbow Method for Alphabet Size = 4 Number of Segment = 25 with usage of Ward linkage

Table 3.4 shows the selected cluster value and "distortion" values of 10, 20, 30 and 40 for each combination and linkage value. According to the obtained results, the best values of the cluster numbers are in bold.

Table 3.4 The Selected Cluster Value and "Distortion" Values Of 10, 20,30 And 40 For Each Combination and Linkage Value

	Alphabet Size=4 Num Segment=25		Alphabet Size=9 Num Segment=25		Alphabet Size=7 Num Segment=180		Alphabet Size=7 Num Segment=45	
	k	Distortion	k	Distortion	k	Distortion	k	Distortion
AVERAGE	**9**	**254.192**	10	781.924	10	4209.415	10	1051.639
	10	245.758	**14**	**576.114**	**13**	**3402.235**	**17**	**622.595**
	20	133.316	20	376.344	20	2688.073	20	571.194
	30	81.297	30	209.324	30	1379.407	30	324.020
	40	41.417	40	135.218	40	903.196	40	198.037
COMPLETE	10	232.802	10	676.567	10	3548.976	10	830.866
	14	**164.017**	**12**	**566.121**	**14**	**2746.310**	**14**	**630.459**
	20	120.780	20	316.535	20	2066.459	20	477.969
	30	71.003	30	190.267	30	1263.865	30	288.717
	40	45.403	40	125.906	40	804.551	40	184.986
SINGLE	10	330.250	10	955.028	10	4914.653	10	1197.694
	16	**247.375**	**12**	**851.935**	**15**	**3799.892**	**17**	**762.054**
	20	223.689	20	559.396	20	2831.060	20	684.257
	30	159.406	30	414.942	30	2137.360	30	515.194
	40	78.667	40	218.514	40	1521.564	40	355.600
WARD	10	198.996	10	589.043	10	3327.717	10	809.397
	13	**161.965**	**12**	**500.989**	**12**	**2943.730**	**12**	**704.709**
	20	110.289	20	298.750	20	1940.199	20	449.367
	30	66.333	30	183.834	30	1226.483	30	278.199
	40	39.233	40	118.250	40	802.700	40	183.369

In Table 3.5, there is an example of companies in clusters are shown for Alphabet Size “4” and Number of Segment “25” combination for Ward linkage with different cluster numbers such as 10, 13. According to the 10 clusters, BRSAN, IZMDC, ULKER belong at the first cluster (C1), EREGL, ISDMR, ISGYO, KRDMD, PETKM, SNGYO, VESBE are in the Cluster 2. At the cluster 4, AEFES, BRYAT, CIMSA, ECZYT, EGEEN, ENJSA, GUBRF, HEKTS, ISMEN, KCHOL, KONYA, KOZAL, MAVI, OTKAR, PGSUS, SASA, THYAO, TKNSA, TOASO, TTRAK, VESTL are assigned. However, AKFGY is only assigned individually. Similarly, according to 13 cluster, BRSAN and ULKER are in the same cluster at the C4. IZDMC is in a different cluster (C11). EREGL, ISDMR, KRDMD, PETKM, VESBE are in the C12. Although ISGYO and SNGYO were previously in this cluster, they are now separated as a separate cluster (C7).

Table 3.5 An Example of Companies in Clusters are shown for Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage

# of Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
10	[BRSAN, IZMDC, ULKER]	[EREGL, ISDMR, ISGYO, KRDMD, PETKM, SNGYO, VESBE]	[BAGFS, GLYHO, HALKB, ODAS, SKBNK, TUKAS, VAKBN, ZOREN]	[AEFES, BRYAT, CIMSA, ECZYT, EGEEN, ENJSA, GUBRF, HEKTS, ISMEN, KCHOL, KONYA, KOZAL, MAVI, OTKAR, PGSUS, SASA, THYAO, TKNSA, TOASO, TTRAK, VESTL]	[ALBRK, BERA, BUCIM]	[AKCNS, ASELS, BIMAS, DOHOL, ECILC, EKGYO, GSDHO, MGROS, OYAKC, SELEC, SOKM, TCELL]	[AGHOL, AKBNK, ALARK, ISCTR, TKFEN, TSKB, TTKOM, YKBNK]	[AKFGY]	[ARCLK, CCOLA, CEMTS, DOAS, FROTO, IPEKE, KARSN, KORDS, KOZAA]	[AKSA, AKSEN, ASUZU, ENKAI, GARAN, SAHOL, SISE, TAVHL, TUPRS]	-	-	-
13	[BAGFS, GLYHO, ODAS, SKBNK, ZOREN]	[AEFES, BRYAT, CIMSA, ECZYT, EGEEN, ENJSA, GUBRF, HEKTS, ISMEN, KCHOL, KONYA, KOZAL, MAVI, OTKAR, PGSUS, SASA, THYAO, TKNSA, TOASO, TTRAK, VESTL]	[AGHOL, AKBNK, ALARK, ISCTR, TKFEN, TSKB, TTKOM, YKBNK]	[BRSAN, ULKER]	[ALBRK, BERA, BUCIM]	[AKCNS, ASELS, BIMAS, DOHOL, ECILC, EKGYO, GSDHO, MGROS, OYAKC, SELEC, SOKM, TCELL]	[ISGYO, SNGYO]	[AKFGY]	[ARCLK, CCOLA, CEMTS, DOAS, FROTO, IPEKE, KARSN, KORDS, KOZAA]	[AKSA, AKSEN, ASUZU, ENKAI, GARAN, SAHOL, SISE, TAVHL, TUPRS]	[IZMDC]	[EREGL, ISDMR, KRDMD, PETKM, VESBE]	[HALKB, TUKAS, VAKBN]

Different clusters for analysing companies' movements for 4 different segments and alphabet numbers. 5 different cluster numbers are analysed in 4 different segment and alphabet number combinations for different linkage methods. A total of 20 different clusters are obtained for each linkage method. It is determined which cluster each company is in. In addition to this, it is aimed to determine which companies are in the same group with each other and how many times they will in same cluster. Table 3.6 shows example of the bilateral movements of the companies are analysed and an 81x81 matrix is created for 81 companies for Ward linkage method. For instance, AKBNK and AEFES moved together in 12 of 20 clusters. Moreover, AKBNK and YKBNK moved together in 15 of 20 clusters. On the other hand, AKBNK could not move together with VESBE or ULKER in any cluster. The results of all other combinations and options are given in Appendix B-E Parts.

Table 3.6 The Example of The Bilateral Total Movements of The Companies

	AEFES	AGHOL	AKBNK	AKCNS	AKSA		TUPRS	VAKBN	VESBE	VESTL	YKBNK	ZOREN
AEFES	0	11	12	0	0		8	0	0	3	9	0
AGHOL	11	0	12	0	0		9	0	0	0	11	0
AKBNK	12	12	0	0	0		8	0	0	0	15	0
AKCNS	0	0	0	0	0		0	0	0	6	0	0
AKFGY	0	0	0	0	0		0	0	0	0	0	0
AKSA	0	0	0	0	0		3	0	2	0	0	0
⋮						...	⋮					
TUPRS	8	9	8	0	3		0	0	0	0	10	0
ULKER	0	0	0	0	0		0	0	0	0	0	0
VAKBN	0	0	0	0	0		0	0	0	0	0	5
VESBE	0	0	0	0	2		0	0	0	0	0	0
VESTL	3	0	0	6	0		0	0	0	0	0	0
YKBNK	9	11	15	0	0		10	0	0	0	0	0
ZOREN	0	0	0	0	0		0	5	0	0	0	0

Chapter 4

Conclusions and Future Prospects

4.1 Conclusions

In this study, the SAX method is used to group 81 stocks selected from the BIST-100 and autonomously identify which stocks have similar characteristics in terms of price movements. We have obtained multiple detailed results based on different combinations of parameters, because of the size limitation however, in this section, we summarize the results and discussions yielded by optimum number of clusters which is 13.

As shown in Table 3.5, in the first scenario (each scenario is a combination of alphabet size and number of segments), the last data collection date for stocks divided into 10 and 13 clusters is 08.08.2023. The number of clusters 13 is also the optimum number of clusters generated by the Ward model. As an example, analysis, when we divide these stocks into 10 clusters, it is seen that one of the clusters is BRSAN, IZDMC, and ULKER stocks. Due to the focus of this study, 3 stocks are expected to exhibit similar behaviour in terms of price movements. When we look at the price movements from 08.08.2023 to 27.11.2023, it is seen that BRSAN and ULKER stocks made approximately 135% and 131% return respectively, while IZDMC stock lost approximately 40% in the same period. This situation is handled by dividing the stocks into 10 different groups and expecting that the stocks in each group will act similarly.

On the other hand, when the shares are divided into 13 different groups for the same 3 stocks, the situation changes completely in parallel with the increase rates seen in Figure 4.1. In this scenario, BRSAN and ULKER are in one group while IZDMC forms a group on its own. This also supports the ward model's determination of the optimum number of clusters as 13.

Figure 4.1 Stock Price movement of BRSAN, ULKER and IZDMC

Based on the focus of the study, the following question can be asked: How can this study help an investor's decision-making process? To present various scenarios based on this question, we have proceeded through the clusters formed with reference to the number of 13 clusters. Our study provides strategic guidance to an investor on the issues listed below.

- Companies acting together in the same sector.

Sector-based movements can be observed in stock market price movements from time to time. These price movements, which may be upward or downward, may cause companies operating in the same sector to exhibit similar movements. In addition, these sector-based movements may not affect the stocks traded in the same sector equally. It will be important to observe which firms in the same sector exhibit similar movements. As one of the results of this study, as shown in Table 3.5, ISGYO and SNGYO stocks are in the same cluster when price movements in the last three years are observed until 08.08.2023. However, when we look at the price movements from 08.08.2023 to 27.11.2023, ISGYO stock provides a return of approximately 23%, while SNGYO stock has a loss of -4%. Therefore, the decision maker can make an inference for SNGYO and ISGYO shares, which should exhibit similar movements with reference to this study but behave differently in a certain period.

- Non-sector companies that can act in parallel with a particular sector.

Another noteworthy point in our study is that while the majority shares in a cluster are from a certain sector, there are companies from other sectors in the same cluster. In other words, non-sectoral firms acting in parallel with a particular sector. When we look at Table 3.5, while Cluster 12 consists mainly of iron and steel shares, PETKIM and VESBE, which are from different sectors, have shown similar movements. A similar situation is valid for Cluster 13; TUKAS, which is in the food sector, showed similar characteristics with HALKB and VAKBN in the banking sector. Investors can look at the historical price movements (last 1 month, last 2 months, etc.) of these stocks, which are expected to be in the same cluster, and shape their investment decisions according to outliers.

- Companies acting independently.

As shown in Table 3.5 and Table 3.6, some firms formed single clusters. For example, only AKFGY and IZDMC shares are included in Clusters 8 and 11, respectively. It may be beneficial for investors who will make investment decisions to consider the fact that such firms may have unique price movements.

- Companies that differ from similar companies negatively/positively

In another scenario, stocks that are positively differentiated from firms in the same cluster - that is, expected to show similar price movements - because of this study. Investors can make more optimized decisions by taking this information into account. For example, as expected, THYAO and PGSUS are in the same cluster as two similar companies in the aviation industry. However, between 08.08.2023 and 27.11.2023, PGSUS suffered a loss of approximately 21% while THYAO suffered a loss of approximately 0.5%. These gain/loss differences in these companies, which are in the same cluster and have similar fields of activity, may help investors to make the right decision.

In addition to these comments, the ideal clusters specified in Table 3.5 are analysed. The closing prices of the companies in each cluster on 08.08.2023 and 27.11.2023 were evaluated. The difference between two dates is calculated as a percentage loss or gain. Additionally, the analysis was made by averaging the percentage earnings of the companies in each cluster. These analyses are shown in Table 4.1.

If we look at the analysis in the clusters, the average return of the companies in the C4 cluster is the highest. On the other hand, while companies in clusters C1, C2, C3, C4, C5,

C6, C7, C10 and C12 made profits, companies in clusters C8, C9 and C11 made losses for investors.

Table 4.1 Net Worth Analysis of Each Cluster

Cluster	Stock	Start (8.08.2023)	End (27.11.2023)	NET WORTH (Percentage)	NET WORTH EACH CLUSTER (Percentage)
C1	BAGFS	32.86	38.74	17.894%	34.301%
	GLYHO	13.40	11.46	-14.478%	
	ODAS	11.16	12.00	7.527%	
	SKBNK	3.19	7.43	132.915%	
	ZOREN	4.16	5.31	27.644%	
C2	AEFES	105.00	115.20	9.714%	13.638%
	BRYAT	1951.40	2380.00	21.964%	
	CIMSA	26.86	35.56	32.404%	
	ECZYT	210.40	326.25	55.062%	
	EGEEN	6437.40	10400.00	61.556%	
	ENJSA	49.12	49.60	0.977%	
	GUBRF	327.00	373.25	14.144%	
	HEKTS	29.70	21.04	-29.158%	
	ISMEN	30.42	33.06	8.679%	
	KCHOL	130.00	142.40	9.538%	
	KONYA	4496.90	9020.00	100.583%	
	KOZAL	28.24	22.82	-19.193%	
	MAVI	93.10	104.60	12.352%	
	OTKAR	322.00	437.25	35.792%	
	PGSUS	923.00	728.00	-21.127%	
	SASA	56.65	46.64	-17.670%	
	THYAO	261.80	260.50	-0.497%	
	TKNSA	26.68	37.18	39.355%	
	TOASO	296.70	241.30	-18.672%	
	TTRAK	787.30	718.50	-8.739%	
	VESTL	60.30	59.90	-0.663%	
C3	AGHOL	163.50	205.30	25.566%	18.414%
	AKBNK	28.16	34.30	21.804%	
	ALARK	110.90	104.20	-6.041%	
	ISCTR	16.23	22.48	38.509%	
	TKFEN	43.80	46.70	6.621%	
	TSKB	6.14	7.59	23.616%	
	TTKOM	20.78	22.84	9.913%	
	YKBNK	14.93	19.01	27.328%	
C4	BRSAN	314.00	740.50	135.828%	122.914%
	ULKER	42.50	89.25	110.000%	
C5	ALBRK	3.33	4.12	23.724%	11.650%
	BERA	14.88	14.50	-2.554%	
	BUCIM	7.91	9.00	13.780%	

C6	AKCNS	116.90	153.10	30.967%	14.470%
	ASELS	38.00	49.40	30.000%	
	BIMAS	231.20	320.50	38.625%	
	DOHOL	12.90	13.62	5.581%	
	ECILC	42.74	49.46	15.723%	
	EKGYO	8.95	7.60	-15.084%	
	GSDHO	4.76	4.15	-12.815%	
	MGROS	331.10	375.75	13.485%	
	OYAKC	62.25	70.00	12.450%	
	SELEC	46.08	52.75	14.475%	
	SOKM	45.26	59.30	31.021%	
	TCELL	53.70	58.65	9.218%	
C7	ISGYO	17.77	21.86	23.016%	9.458%
	SNGYO	3.17	3.04	-4.101%	
C8	AKFGY	4.90	4.68	-4.490%	-4.490%
C9	ARCLK	148.00	142.90	-3.446%	-2.853%
	CCOLA	335.70	415.25	23.697%	
	CEMTS	11.08	11.81	6.588%	
	DOAS	269.50	252.00	-6.494%	
	FROTO	943.80	822.50	-12.852%	
	IPEKE	45.72	41.00	-10.324%	
	KARSN	11.60	9.60	-17.241%	
	KORDS	84.20	84.75	0.653%	
	KOZAA	59.90	56.15	-6.260%	
C10	AKSA	87.60	90.95	3.824%	8.088%
	AKSEN	38.34	35.20	-8.190%	
	ASUZU	255.50	233.50	-8.611%	
	ENKAI	32.94	34.60	5.039%	
	GARAN	48.22	54.75	13.542%	
	SAHOL	57.60	62.55	8.594%	
	SISE	51.45	50.00	-2.818%	
	TAVHL	112.60	121.80	8.171%	
	TUPRS	101.80	156.00	53.242%	
C11	IZMDC	8.60	6.11	-28.953%	-28.953%
C12	EREGL	40.98	41.28	0.732%	5.298%
	ISDMR	39.02	36.20	-7.227%	
	KRDMD	23.60	25.10	6.356%	
	PETKM	18.32	22.26	21.507%	
	VESBE	16.20	17.03	5.123%	
C13	HALKB	13.77	13.35	-3.050%	1.599%
	TUKAS	9.60	8.15	-15.104%	
	VAKBN	12.33	15.16	22.952%	

These scenarios are just a few examples of scenarios and inferences that can help investors make the right decisions because of our study and embody the basic logic of our study.

4.2 Societal Impact and Contribution to Global Sustainability

The study outlined in the abstract holds significant potential for both social impact and global contribution within the realm of financial markets. By employing computational finance techniques like the Symbolic Aggregate Approximation (SAX) method and clustering algorithms, the research aims to democratize access to predictive insights in stock market behaviour. In doing so, it empowers a broader range of stakeholders, including traders and investors, to make informed decisions, thereby potentially reducing information asymmetry and enhancing market efficiency. This democratization of predictive analytics could lead to more equitable participation in financial markets, fostering economic inclusion and potentially mitigating disparities in wealth distribution. Furthermore, the study's focus on the BIST100 companies demonstrates a localized approach that could have ripple effects on the global stage. By providing actionable insights tailored to a specific market, the research contributes to the robustness and stability of financial systems, thereby fostering confidence among domestic and international investors. Moreover, the use of the dendrogram tree graph for analysing clustering patterns adds transparency to the process, enhancing trust and understanding among stakeholders. Overall, the study not only has the potential to revolutionize decision-making in financial markets but also to promote broader societal benefits by levelling the playing field and fostering confidence in investment opportunities, thereby contributing to global economic stability and prosperity.

4.3 Future Prospects

In our pursuit of future studies, we aim to enhance the scope of our current research, which focused on 81 firms selected by segregating newly publicly offered shares from

the BIST100 stocks. Our envisioned expansions include broadening the investigation to encompass BISTTUM, allowing for a more comprehensive understanding of market dynamics. Additionally, we intend to delve deeper into sector-specific analyses, recognizing the nuances and unique factors influencing various industries. Another avenue of exploration involves conducting a parallel study specifically tailored to newly publicly offered companies, enabling a comparative analysis of their performance and market behaviour. Moreover, we seek to refine our approach by customizing the study parameters to companies exhibiting similar technical metrics, such as the P/E ratio and the MV/BV ratio, facilitating a more nuanced evaluation within homogeneous subsets. Furthermore, we aspire to augment the sophistication of our clustering algorithm by incorporating additional attributes such as trading volume, market capitalization, and the number of outstanding shares. By integrating these multifaceted variables, we anticipate a more nuanced and comprehensive understanding of market dynamics, enabling us to discern subtle patterns and trends that might elude conventional analysis. In essence, our future endeavours aim to elevate the depth and breadth of our research, fostering insights that are not only relevant but also actionable in navigating the complexities of the financial landscape.

BIBLIOGRAPHY

[1] O. Uluyol and V. E. Türk, "FİNANSAL RASYOLARIN FİRMA DEĞERİNE ETKİSİ: BORSA İSTANBUL (BİST)'DA BİR UYGULAMA," *Afyon Kocatepe Üniversitesi, İİBF Dergisi*, vol. 15, no. 2, pp. 365–384, Dec. 2013.

[2] Ö. YÜCEL, "FİNANSAL PİYASA ETKİNLİĞİ: BORSA İSTANBUL ÜZERİNE BİR UYGULAMA," *International Review of Economics and Management*, vol. 4, no. 3, Dec. 2016, doi: 10.18825/irem.16916.

[3] A. Kılıç, B. Güloğlu, A. Yalçın, and A. Üstündağ, "Big data–enabled sign prediction for Borsa Istanbul intraday equity prices," *Borsa Istanbul Review*, vol. 23, pp. S38–S52, Dec. 2023, doi: 10.1016/j.bir.2023.08.005.

[4] D. Sheth and M. Shah, "Predicting stock market using machine learning: best and accurate way to know future stock prices," *International Journal of System Assurance Engineering and Management*, vol. 14, no. 1, pp. 1–18, Feb. 2023, doi: 10.1007/s13198-022-01811-1.

[5] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering," *ACM Comput Surv*, vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: 10.1145/331499.331504.

[6] G. J. Oyewole and G. A. Thopil, "Data clustering: application and trends," *Artif Intell Rev*, vol. 56, no. 7, pp. 6439–6475, Jul. 2023, doi: 10.1007/s10462-022-10325-y.

[7] M. Nazari, A. Hussain, and P. Musilek, "Applications of Clustering Methods for Different Aspects of Electric Vehicles," *Electronics (Basel)*, vol. 12, no. 4, p. 790, Feb. 2023, doi: 10.3390/electronics12040790.

[8] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[9] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, New York, NY, USA: ACM, Jun. 2003, pp. 2–11. doi: 10.1145/882082.882086.

[10] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time series," *Data Min Knowl Discov*, vol. 15, no. 2, pp. 107–144, Aug. 2007, doi: 10.1007/s10618-007-0064-z.

[11] F. Murtagh, "A Survey of Recent Advances in Hierarchical Clustering Algorithms," *Comput J*, vol. 26, no. 4, pp. 354–359, Nov. 1983, doi: 10.1093/comjnl/26.4.354.

[12] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, Jan. 2012, doi: 10.1002/widm.53.

[13] S. K. Popat and M. Emmanuel, "Review and comparative study of clustering techniques," *International journal of computer science and information technologies*, vol. 5, no. 1, pp. 805–812, 2014.

[14] J. B. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," *In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[15] D. P. Ismi and M. Murinto, “Clustering based feature selection using Partitioning Around Medoids (PAM),” *Jurnal Informatika*, vol. 14, no. 2, p. 50, May 2020, doi: 10.26555/jifo.v14i2.a17620.
[16] L. J. Hamilton, “Characterising spectral sea wave conditions with statistical clustering of actual spectra,” *Applied Ocean Research*, vol. 32, no. 3, pp. 332–342, Jul. 2010, doi: 10.1016/j.apor.2009.12.003.
[17] S. Vijayarani and S. Nithya, “An Efficient Clustering Algorithm For Outlier Detection,” *Int J Comput Appl*, vol. 32, no. 7, pp. 22–27, 2011.
[18] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, “Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases,” *Knowl Inf Syst*, vol. 3, no. 3, pp. 263–286, Aug. 2001, doi: 10.1007/PL00011669.
[19] L. van der Maaten, E. Postma, and J. van den Herik, “Dimensionality Reduction: A Comparative Review,” *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.
[20] H. Ren, X. Liao, Z. Li, and A. AI-Ahmari, “Anomaly detection using piecewise aggregate approximation in the amplitude domain,” *Applied Intelligence*, vol. 48, no. 5, pp. 1097–1110, May 2018, doi: 10.1007/s10489-017-1017-x.
[21] M. De Oliveira Jr, G. Sedrez, and G. G. H. Cavalheiro, “ML-based Plant Stress Detection from IoT-sensed Reduced Electromes,” *The International FLAIRS Conference Proceedings*, vol. 36, May 2023, doi: 10.32473/flairs.36.133180.
[22] Y. Yu, T. Becker, L. M. Trinh, and M. Behrisch, “SAXRegEx: Multivariate time series pattern search with symbolic representation, regular expression, and query expansion,” *Comput Graph*, vol. 112, pp. 13–21, May 2023, doi: 10.1016/j.cag.2023.03.002.
[23] S. Madhulatha, “AN OVERVIEW ON CLUSTERING METHODS,” *IOSR Journal of Engineering*, vol. 2, no. 4, pp. 719–725, Apr. 2012.
[24] L. Zhang *et al.*, “Recognition of oil & gas pipelines operational states using graph network structural features,” *Eng Appl Artif Intell*, vol. 120, p. 105884, Apr. 2023, doi: 10.1016/j.engappai.2023.105884.
[25] M. Barbar and D. Mallapragada, “). Representative period selection for power system planning using autoencoder-based dimensionality reduction.,” 2022.
[26] W.-K. Jung, H. Kim, Y.-C. Park, J.-W. Lee, and E. S. Suh, “Real-time data-driven discrete-event simulation for garment production lines,” *Production Planning & Control*, vol. 33, no. 5, pp. 480–491, Apr. 2022, doi: 10.1080/09537287.2020.1830194.
[27] B. G. Sürmeli and M. B. Tümer, “Multivariate Time Series Clustering and its Application in Industrial Systems,” *Cybern Syst*, vol. 51, no. 3, pp. 315–334, Apr. 2020, doi: 10.1080/01969722.2019.1691851.
[28] K. Villalobos, B. Diez, A. Illarramendi, A. Goñi, and J. M. Blanco, “I4TSRS: A System to Assist a Data Engineer in Time-Series Dimensionality Reduction in Industry 4.0 Scenarios,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, New York, NY, USA: ACM, Oct. 2018, pp. 1915–1918. doi: 10.1145/3269206.3269213.
[29] M. Van Hoan, D. T. Huy, and L. C. Mai, “Pattern Discovery in the Financial Time Series Based on Local Trend,” 2017, pp. 442–451. doi: 10.1007/978-3-319-49073-1_48.
[30] W. Liu and L. Shao, “Research of SAX in Distance Measuring for Financial Time Series Data,” in *2009 First International Conference on Information Science and Engineering*, IEEE, 2009, pp. 935–937. doi: 10.1109/ICISE.2009.924.

[31] C.-H. Chen and C.-H. Yu, “A Series-based group stock portfolio optimization approach using the grouping genetic algorithm with symbolic aggregate Approximations,” *Knowl Based Syst*, vol. 125, pp. 146–163, Jun. 2017, doi: 10.1016/j.knosys.2017.03.018.

[32] A. Roques and A. Zhao, “Association Rules Discovery of Deviant Events in Multivariate Time Series: An Analysis and Implementation of the SAX-ARM Algorithm,” *Image Processing On Line*, vol. 12, pp. 604–624, Dec. 2022, doi: 10.5201/ipol.2022.437.

[33] H. Park and J.-Y. Jung, “SAX-ARM: Deviant event pattern discovery from multivariate time series using symbolic aggregate approximation and association rule mining,” *Expert Syst Appl*, vol. 141, p. 112950, Mar. 2020, doi: 10.1016/j.eswa.2019.112950.

[34] M. M. Muhammad Fuad, “Extreme-SAX: extreme points based symbolic representation for time series classification.,” in *In Big Data Analytics and Knowledge Discovery: 22nd International Conference, DaWaK 2020, Bratislava, Slovakia, September 14–17*, Springer International Publishing., 2020, pp. 122–130.

[35] P. M. Chau, B. M. Duc, and D. T. Anh, “Discord Discovery in Streaming Time Series based on an Improved HOT SAX Algorithm,” in *Proceedings of the Ninth International Symposium on Information and Communication Technology - SoICT 2018*, New York, New York, USA: ACM Press, 2018, pp. 24–30. doi: 10.1145/3287921.3287929.

[36] E. Keogh, J. Lin, and A. Fu, “HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence,” in *Fifth IEEE International Conference on Data Mining (ICDM’05)*, IEEE, pp. 226–233. doi: 10.1109/ICDM.2005.79.

[37] N. D. Pham, Q. L. Le, and T. K. Dang, “HOT aSAX: A Novel Adaptive Symbolic Representation for Time Series Discords Discovery,” 2010, pp. 113–121. doi: 10.1007/978-3-642-12145-6_12.

[38] P. Avogadro and M. A. Dominoni, “A fast algorithm for complex discord searches in time series: HOT SAX Time,” *Applied Intelligence*, vol. 52, no. 9, pp. 10060–10081, Jul. 2022, doi: 10.1007/s10489-021-02897-z.

[39] B. Lkhagva, Y. Suzuki, and K. Kawagoe, “Extended SAX: Extension of Symbolic Aggregate Approximation for Financial Time Series Data Representation,” *DEWS2006 4A-i8* .

[40] C. T. Zan and H. Yamana, “Dynamic SAX parameter estimation for time series,” *International Journal of Web Information Systems*, vol. 13, no. 4, pp. 387–404, Nov. 2017, doi: 10.1108/IJWIS-04-2017-0035.

[41] P. Senin and S. Malinchik, “SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model,” in *2013 IEEE 13th International Conference on Data Mining*, IEEE, Dec. 2013, pp. 1175–1180. doi: 10.1109/ICDM.2013.52.

[42] Y. Sun, J. Li, J. Liu, B. Sun, and C. Chow, “An improvement of symbolic aggregate approximation distance measure for time series,” *Neurocomputing*, vol. 138, pp. 189–198, Aug. 2014, doi: 10.1016/j.neucom.2014.01.045.

[43] S. Malinowski, T. Guyet, R. Quiniou, and R. Tavenard, “1d-SAX: A Novel Symbolic Representation for Time Series,” *ADVANCES IN INTELLIGENT DATA ANALYSIS XII*, vol. 8207, pp. 273–284, Dec. 2013.

[44] B. Bai, G. Li, S. Wang, Z. Wu, and W. Yan, “Time series classification based on multi-feature dictionary representation and ensemble learning,” *Expert Syst Appl*, vol. 169, p. 114162, May 2021, doi: 10.1016/j.eswa.2020.114162.

[45] Z. He, C. Zhang, and Y. Cheng, “Similarity Measurement and Classification of Temporal Data Based on Double Mean Representation,” *Algorithms*, vol. 16, no. 7, p. 347, Jul. 2023, doi: 10.3390/a16070347.
[46] D.-H. Yang and Y.-S. Kang, “Distance- and Momentum-Based Symbolic Aggregate Approximation for Highly Imbalanced Classification,” *Sensors*, vol. 22, no. 14, p. 5095, Jul. 2022, doi: 10.3390/s22145095.
[47] J. Yang, S. Jing, and G. Huang, “Accurate and fast time series classification based on compressed random Shapelet Forest,” *Applied Intelligence*, Jun. 2022, doi: 10.1007/s10489-022-03852-2.
[48] G. Li, B. Choi, J. Xu, S. S. Bhowmick, K.-P. Chun, and G. L.-H. Wong, “Efficient Shapelet Discovery for Time Series Classification,” *IEEE Trans Knowl Data Eng*, vol. 34, no. 3, pp. 1149–1163, Mar. 2022, doi: 10.1109/TKDE.2020.2995870.
[49] R. J. Kate, “Using dynamic time warping distances as features for improved time series classification,” *Data Min Knowl Discov*, vol. 30, no. 2, pp. 283–312, Mar. 2016, doi: 10.1007/s10618-015-0418-x.
[50] Z. He, C. Zhang, X. Ma, and G. Liu, “Hexadecimal Aggregate Approximation Representation and Classification of Time Series Data,” *Algorithms*, vol. 14, no. 12, p. 353, Dec. 2021, doi: 10.3390/a14120353.
[51] J. Lin *et al.*, “Predictive analytics for building power demand: Day-ahead forecasting and anomaly prediction,” *Energy Build*, vol. 255, p. 111670, Jan. 2022, doi: 10.1016/j.enbuild.2021.111670.
[52] N. T. N. Anh, P. N. Q. Anh, V. H. Thu, D. Van Thai, V. K. Solanki, and D. M. Tuan, “A novel approach for anomaly detection in automatic meter intelligence system using machine learning and pattern recognition,” *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 1843–1852, Jun. 2022, doi: 10.3233/JIFS-219285.
[53] R. Moskovitch and Y. Shahar, “Classification of multivariate time series via temporal abstraction and time intervals mining,” *Knowl Inf Syst*, vol. 45, no. 1, pp. 35–74, Oct. 2015, doi: 10.1007/s10115-014-0784-5.
[54] S. Cohen, O. Katz, D. Presil, O. Arbili, and L. Rokach, “Ensemble Learning for Alcoholism Classification Using EEG Signals,” *IEEE Sens J*, vol. 23, no. 15, pp. 17714–17724, Aug. 2023, doi: 10.1109/JSEN.2023.3279904.
[55] N. Tabassum, S. Menon, and A. Jastrzębska, “Time-series classification with SAFE: Simple and fast segmented word embedding-based neural time series classifier,” *Inf Process Manag*, vol. 59, no. 5, p. 103044, Sep. 2022, doi: 10.1016/j.ipm.2022.103044.
[56] Y. R. Veeranki, N. Ganapathy, and R. Swaminathan, “Analysis of Fluctuation Patterns in Emotional States Using Electrodermal Activity Signals and Improved Symbolic Aggregate Approximation,” *Fluctuation and Noise Letters*, vol. 21, no. 02, Apr. 2022, doi: 10.1142/S0219477522500134.
[57] Y. Wan, X. Gong, and Y.-W. Si, “Effect of segmentation on financial time series pattern matching,” *Appl Soft Comput*, vol. 38, pp. 346–359, Jan. 2016, doi: 10.1016/j.asoc.2015.10.012.
[58] F. Delbianco, A. Fioriti, and F. Tohmé, “Designing a contemporaneity index: Detecting regional similarities in South America, 1961-2018,” *Regional Statistics*, vol. 13, no. 4, pp. 634–650, 2023, doi: 10.15196/RS130403.
[59] J. Liu, W. Huang, H. Li, S. Ji, Y. Du, and T. Li, “SLAFusion: Attention fusion based on SAX and LSTM for dangerous driving behavior detection,” *Inf Sci (N Y)*, vol. 640, p. 119063, Sep. 2023, doi: 10.1016/j.ins.2023.119063.

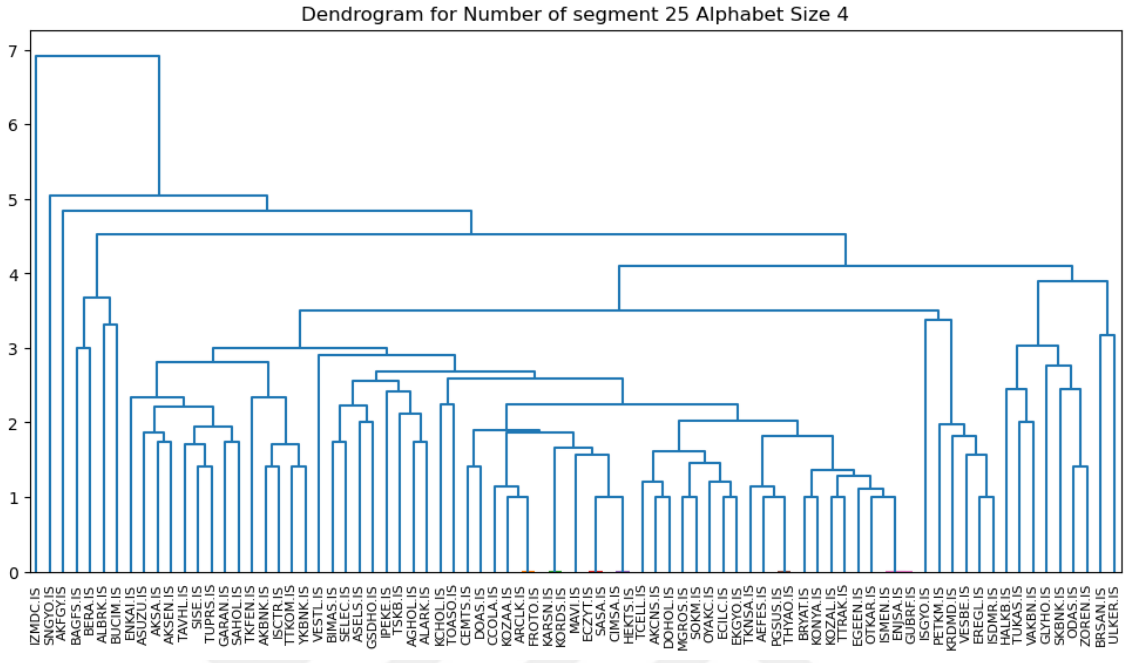
[60] J. Liu, T. Li, Z. Yuan, W. Huang, P. Xie, and Q. Huang, “Symbolic aggregate approximation based data fusion model for dangerous driving behavior detection,” *Inf Sci (N Y)*, vol. 609, pp. 626–643, Sep. 2022, doi: 10.1016/j.ins.2022.07.118.
[61] P. T. Alluri, H. Banna, M. R. Khalghani, S. Khushalani-Solanki, and J. Solanki, “Real-Time Framework for Monitoring Cyber Disruptions in Power Grids,” *IEEE Trans Industr Inform*, vol. 18, no. 6, pp. 4008–4017, Jun. 2022, doi: 10.1109/TII.2021.3105679.
[62] D. Zhao, S. Liu, Z. Miao, H. Zhang, Y. Wei, and S. Xiao, “A Novel Feature Extraction Approach for Mechanical Fault Diagnosis Based on ESAX and BoW Model,” *IEEE Trans Instrum Meas*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3185658.
[63] B. Wang, Y. Ning, and Y. Zhang, “A novel fault diagnosis scheme for rolling bearing based on symbolic aggregate approximation and convolutional neural network with channel attention,” *Meas Sci Technol*, vol. 33, no. 1, p. 015016, Jan. 2022, doi: 10.1088/1361-6501/ac319a.
[64] G. Li, M. Yin, S. Jing, and B. Guo, “An Effective Algorithm for Intrusion Detection Using Random Shapelet Forest,” *Wirel Commun Mob Comput*, vol. 2021, pp. 1–9, Sep. 2021, doi: 10.1155/2021/4214784.
[65] Z. Wang, L. Wang, C. Huang, Z. Zhang, and X. Luo, “Soil-Moisture-Sensor-Based Automated Soil Water Content Cycle Classification With a Hybrid Symbolic Aggregate Approximation Algorithm,” *IEEE Internet Things J*, vol. 8, no. 18, pp. 14003–14012, Sep. 2021, doi: 10.1109/JIOT.2021.3068379.
[66] S. Guo and W. Guo, “Process Monitoring and Fault Prediction in Multivariate Time Series Using Bag-of-Words,” *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 1, pp. 230–242, Jan. 2022, doi: 10.1109/TASE.2020.3026065.
[67] X. Chen, J. Peng, Z. Song, Y. Zheng, and B. Zhang, “Monitoring Persistent Coal Fire Using Landsat Time Series Data From 1986 to 2020,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2022.3142350.
[68] J. Earnest, “Sum of Gaussian Feature-Based Symbolic Representations of Eddy Current Defect Signatures,” *Research in Nondestructive Evaluation*, vol. 34, no. 3–4, pp. 136–153, Jul. 2023, doi: 10.1080/09349847.2023.2217094.
[69] J. Liu, X. Guo, S. Xu, and Y. Zhang, “Quantifying the impact of strong ties in international scientific research collaboration,” *PLoS One*, vol. 18, no. 1, p. e0280521, Jan. 2023, doi: 10.1371/journal.pone.0280521.
[70] B. Esmael, A. Arnaout, R. K. Fruhwirth, and G. Thonhauser, “Multivariate Time Series Classification by Combining Trend-Based and Value-Based Approximations,” 2012, pp. 392–403. doi: 10.1007/978-3-642-31128-4_29.
[71] G. Georgoulas, P. Karvelis, T. Loutas, and C. D. Stylios, “Rolling element bearings diagnostics using the Symbolic Aggregate approXimation,” *Mech Syst Signal Process*, vol. 60–61, pp. 229–242, Aug. 2015, doi: 10.1016/j.ymssp.2015.01.033.
[72] Y. Wang, Q. Chen, C. Kang, and Q. Xia, “Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications,” *IEEE Trans Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016, doi: 10.1109/TSG.2016.2548565.
[73] S. Choi and S. Yoon, “Energy signature-based clustering using open data for urban building energy analysis toward carbon neutrality: A case study on

electricity change under COVID-19," *Sustain Cities Soc*, vol. 92, p. 104471, May 2023, doi: 10.1016/j.scs.2023.104471.
[74] Y. Hong, S. Yoon, and S. Choi, "Operational signature-based symbolic hierarchical clustering for building energy, operation, and efficiency towards carbon neutrality," *Energy*, vol. 265, p. 126276, Feb. 2023, doi: 10.1016/j.energy.2022.126276.
[75] G. Stilo and P. Velardi, "Efficient temporal mining of micro-blog texts and its application to event discovery," *Data Min Knowl Discov*, vol. 30, no. 2, pp. 372–402, Mar. 2016, doi: 10.1007/s10618-015-0412-3.
[76] G. You, "Urban Mobility and Knowledge Extraction from Chaotic Time Series Data: A Comparative Analysis for Uncovering COVID-19 Effects," *Ann Am Assoc Geogr*, vol. 113, no. 9, pp. 2166–2185, Oct. 2023, doi: 10.1080/24694452.2023.2216773.
[77] H. Huang, R. Zhang, C. Xie, and X. Li, "Identifying Subway Passenger Flow under Large-Scale Events Using Symbolic Aggregate Approximation Algorithm," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2676, no. 2, pp. 800–810, Feb. 2022, doi: 10.1177/03611981211047835.
[78] S. Jeon, B. Hong, and V. Chang, "Pattern graph tracking-based stock price prediction using big data," *Future Generation Computer Systems*, vol. 80, pp. 171–187, Mar. 2018, doi: 10.1016/j.future.2017.02.010.
[79] A. M. Hussein, A. K. Idrees, and R. Couturier, "A distributed prediction–compression-based mechanism for energy saving in IoT networks," *J Supercomput*, vol. 79, no. 15, pp. 16963–16999, Oct. 2023, doi: 10.1007/s11227-023-05317-w.
[80] A. M. Hussein, A. K. Idrees, and R. Couturier, "Distributed energy-efficient data reduction approach based on prediction and compression to reduce data transmission in IoT networks," *International Journal of Communication Systems*, vol. 35, no. 15, Oct. 2022, doi: 10.1002/dac.5282.
[81] T. Zhou, Y. Wang, Q. Zhu, and J. Du, "Human hand motion prediction based on feature grouping and deep learning: Pipe skid maintenance example," *Autom Constr*, vol. 138, p. 104232, Jun. 2022, doi: 10.1016/j.autcon.2022.104232.
[82] B. Lei, P. Zhang, Y. Suo, and N. Li, "SAX-STGCN: Dynamic Spatio-Temporal Graph Convolutional Networks for Traffic Flow Prediction," *IEEE Access*, vol. 10, pp. 107022–107031, 2022, doi: 10.1109/ACCESS.2022.3211518.
[83] A. E. Tozzi *et al.*, "Digital Surveillance Through an Online Decision Support Tool for COVID-19 Over One Year of the Pandemic in Italy: Observational Study," *J Med Internet Res*, vol. 23, no. 8, p. e29556, Aug. 2021, doi: 10.2196/29556.
[84] G. Ahn, H. Yun, S. Hur, and S. Lim, "A Time-Series Data Generation Method to Predict Remaining Useful Life," *Processes*, vol. 9, no. 7, p. 1115, Jun. 2021, doi: 10.3390/pr9071115.
[85] J. Yin, Y. Li, R. Wang, and M. Xu, "An Improved Similarity Trajectory Method Based on Monitoring Data under Multiple Operating Conditions," *Applied Sciences*, vol. 11, no. 22, p. 10968, Nov. 2021, doi: 10.3390/app112210968.
[86] Z.-C. Wen *et al.*, "Tri-Partition Alphabet-Based State Prediction for Multivariate Time-Series," *Applied Sciences*, vol. 11, no. 23, p. 11294, Nov. 2021, doi: 10.3390/app112311294.

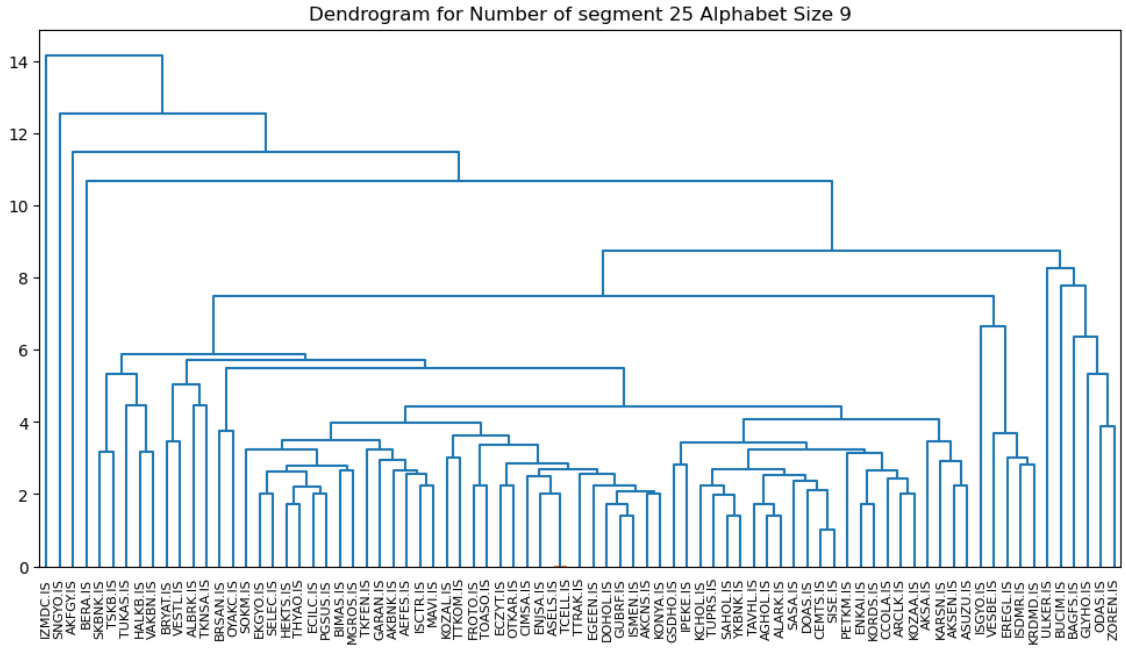
[87] C. Miller, Z. Nagy, and A. Schlueter, “Automated daily pattern filtering of measured building performance data,” *Autom Constr*, vol. 49, pp. 1–17, Jan. 2015, doi: 10.1016/j.autcon.2014.09.004.
[88] D. Popescu, F. Stoican, G. Stamatescu, L. Ichim, and C. Dragana, “Advanced UAV–WSN System for Intelligent Monitoring in Precision Agriculture,” *Sensors*, vol. 20, no. 3, p. 817, Feb. 2020, doi: 10.3390/s20030817.
[89] C. Fan, F. Xiao, H. Madsen, and D. Wang, “Temporal knowledge discovery in big BAS data for building energy management,” *Energy Build*, vol. 109, pp. 75–89, Dec. 2015, doi: 10.1016/j.enbuild.2015.09.060.
[90] A. Capozzoli, M. S. Piscitelli, S. Brandi, D. Grassi, and G. Chicco, “Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings,” *Energy*, vol. 157, pp. 336–352, Aug. 2018, doi: 10.1016/j.energy.2018.05.127.
[91] M. J. Afroni, D. Sutanto, and D. Stirling, “Analysis of Nonstationary Power-Quality Waveforms Using Iterative Hilbert Huang Transform and SAX Algorithm,” *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2134–2144, Oct. 2013, doi: 10.1109/TPWRD.2013.2264948.
[92] S. Kolozali *et al.*, “Observing the Pulse of a City: A Smart City Framework for Real-Time Discovery, Federation, and Aggregation of Data Streams,” *IEEE Internet Things J*, vol. 6, no. 2, pp. 2651–2668, Apr. 2019, doi: 10.1109/JIOT.2018.2872606.
[93] A. Mueen, “Time series motif discovery: dimensions and applications,” *WIREs Data Mining and Knowledge Discovery*, vol. 4, no. 2, pp. 152–159, Mar. 2014, doi: 10.1002/widm.1119.
[94] L. Mascali *et al.*, “A machine learning-based Anomaly Detection Framework for building electricity consumption data,” *Sustainable Energy, Grids and Networks*, vol. 36, p. 101194, Dec. 2023, doi: 10.1016/j.segan.2023.101194.
[95] T. N. A. Nguyen, H. T. Vu, M. T. Dang, D. Kim, and A. N. Le, “Anomaly Detection in Automatic Meter Intelligence System Using Positive Unlabeled Learning and Multiple Symbolic Aggregate Approximation,” *Big Data*, vol. 11, no. 3, pp. 225–238, Jun. 2023, doi: 10.1089/big.2021.0471.
[96] M. Wang, Q. Ge, C. Li, and C. Sun, “Charging Diagnosis of Power Battery Based on Adaptive STCKF and BLS for Electric Vehicles,” *IEEE Trans Veh Technol*, vol. 71, no. 8, pp. 8251–8265, Aug. 2022, doi: 10.1109/TVT.2022.3171766.
[97] L. S. Riza, M. N. Fazanadi, J. A. Utama, K. A. F. A. Samah, T. Hidayat, and S. Nazir, “SAX and Random Projection Algorithms for the Motif Discovery of Orbital Asteroid Resonance Using Big Data Platforms,” *Sensors*, vol. 22, no. 14, p. 5071, Jul. 2022, doi: 10.3390/s22145071.
[98] A. G. Martín, I. Martín de Diego, A. Fernández-Isabel, M. Beltrán, and R. R. Fernández, “Combining user behavioural information at the feature level to enhance continuous authentication systems,” *Knowl Based Syst*, vol. 244, p. 108544, May 2022, doi: 10.1016/j.knosys.2022.108544.
[99] E. Cartwright, M. Crane, and H. J. Ruskin, “Side-Length-Independent Motif (SLIM): Motif Discovery and Volatility Analysis in Time Series—SAX, MDL and the Matrix Profile,” *Forecasting*, vol. 4, no. 1, pp. 219–237, Feb. 2022, doi: 10.3390/forecast4010013.
[100] Lala Septem Riza, Muhammad Naufal Fazanadi, Judhistira Aria Utama, Taufiq Hidayat, and Khyrina Airin Fariza Abu Samah, “The implementation of sax and random projection for motif discovery on the orbital elements and the resonance

argument of asteroid," *International Journal of Nonlinear Analysis and Applications*, vol. 12, no. Special Issue, pp. 959–970, Jun. 2022.
[101] A. Notaristefano, G. Chicco, and F. Piglione, "Data size reduction with symbolic aggregate approximation for electrical load pattern grouping," *IET Generation, Transmission & Distribution*, vol. 7, no. 2, pp. 108–117, Feb. 2013, doi: 10.1049/iet-gtd.2012.0383.
[102] Y. Liu *et al.*, "Knowledge Discovery and Diagnosis Using Temporal-Association-Rule-Mining-Based Approach for Threshing Cylinder Blockage," *Agriculture*, vol. 13, no. 7, p. 1299, Jun. 2023, doi: 10.3390/agriculture13071299.
[103] Merkezi Kayıt İstanbul, "Public Disclosure Platform."
[104] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl Soft Comput*, vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.
[105] E. Ogasawara, L. C. Martinez, D. de Oliveira, G. Zimbrao, G. L. Pap, and M. Mattoso, "Adaptive Normalization: A novel data normalization approach for non-stationary time series," in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2010, pp. 1–8. doi: 10.1109/IJCNN.2010.5596746.
[106] M. G. Baydogan and G. Runger, "Time series representation and similarity based on local autopatterns," *Data Min Knowl Discov*, vol. 30, no. 2, pp. 476–509, Mar. 2016, doi: 10.1007/s10618-015-0425-y.
[107] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco, "A time series representation model for accurate and fast similarity detection," *Pattern Recognit*, vol. 42, no. 11, pp. 2998–3014, Nov. 2009, doi: 10.1016/j.patcog.2009.03.030.
[108] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases," *Knowl Inf Syst*, vol. 3, no. 3, pp. 263–286, Aug. 2001, doi: 10.1007/PL00011669.
[109] H. Yahyaoui, H. AboElfotoh, and Y. Shu, "A multilevel adaptive reduction technique for time series," *Neurocomputing*, vol. 555, p. 126657, Oct. 2023, doi: 10.1016/j.neucom.2023.126657.
[110] N. I. Boyko and O. A. Tkachyk, "Hierarchical clustering algorithm for dendrogram construction and cluster counting," *Informatics and mathematical methods in simulation*, vol. 13, no. 1–2, pp. 5–15, Apr. 2023, doi: 10.15276/imms.v13.no1-2.5.
[111] O. Yim and K. T. Ramdeen, "Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data," *Quant Method Psychol*, vol. 11, no. 1, pp. 8–21, Feb. 2015, doi: 10.20982/tqmp.11.1.p008.
[112] Vijaya, S. Sharma, and N. Batra, "Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, IEEE, Feb. 2019, pp. 568–573. doi: 10.1109/COMITCon.2019.8862232.
[113] Purnima Bholowalia and Arvind Kumar, "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN," *Int J Comput Appl*, vol. 105, no. 9, pp. 17–24, Sep. 2014.
[114] Mengyao Cui, "Introduction to the K-Means Clustering Algorithm Based on the Elbow Method," *Accounting, Auditing and Finance*, vol. 3, pp. 9–16, 2020.

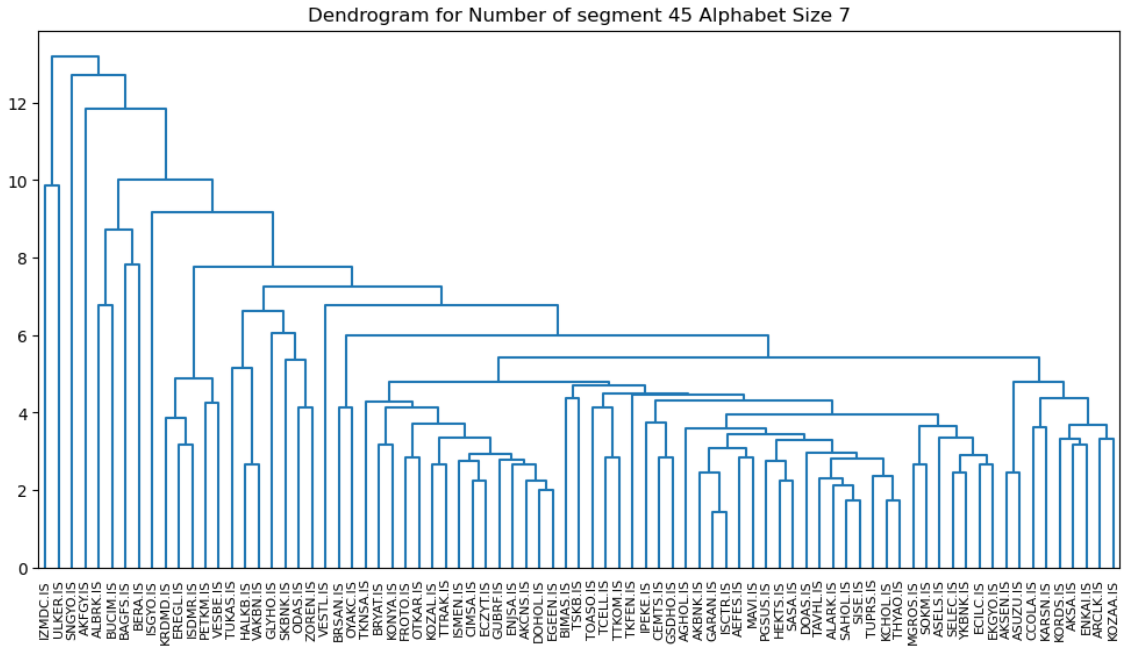
APPENDIX



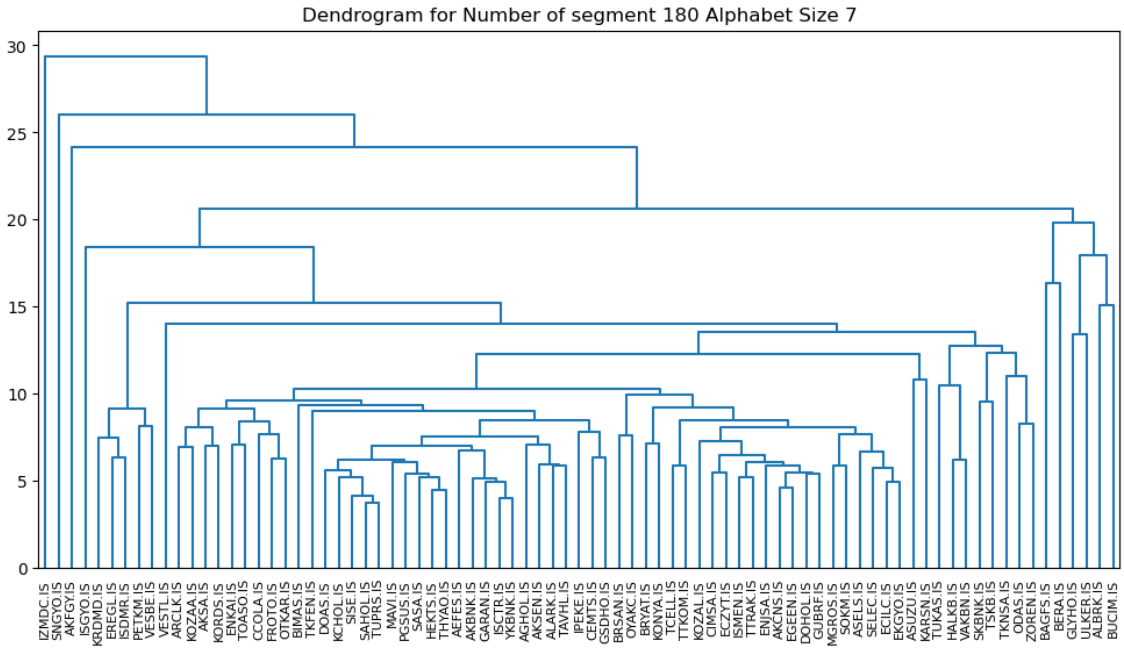
Appendix A 1 Dendrogram for Alphabet Size = 4 Number of Segment = 25 With Usage of Average Linkage



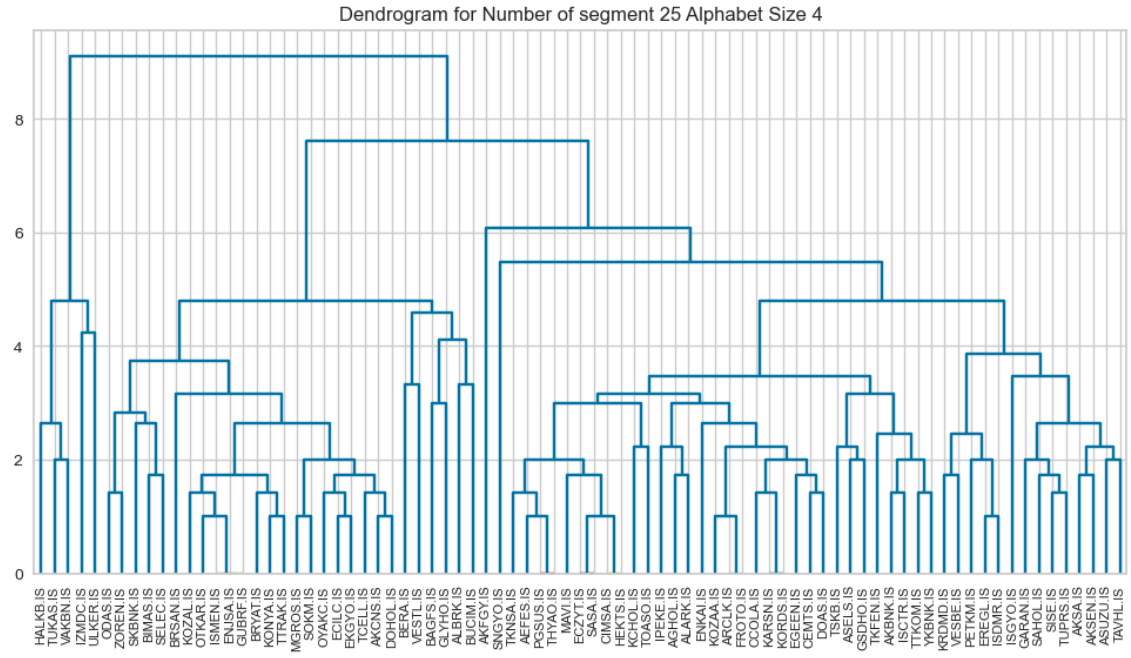
Appendix A 2 Dendrogram for Alphabet Size = 9 Number of Segment = 25 With Usage of Average Linkage



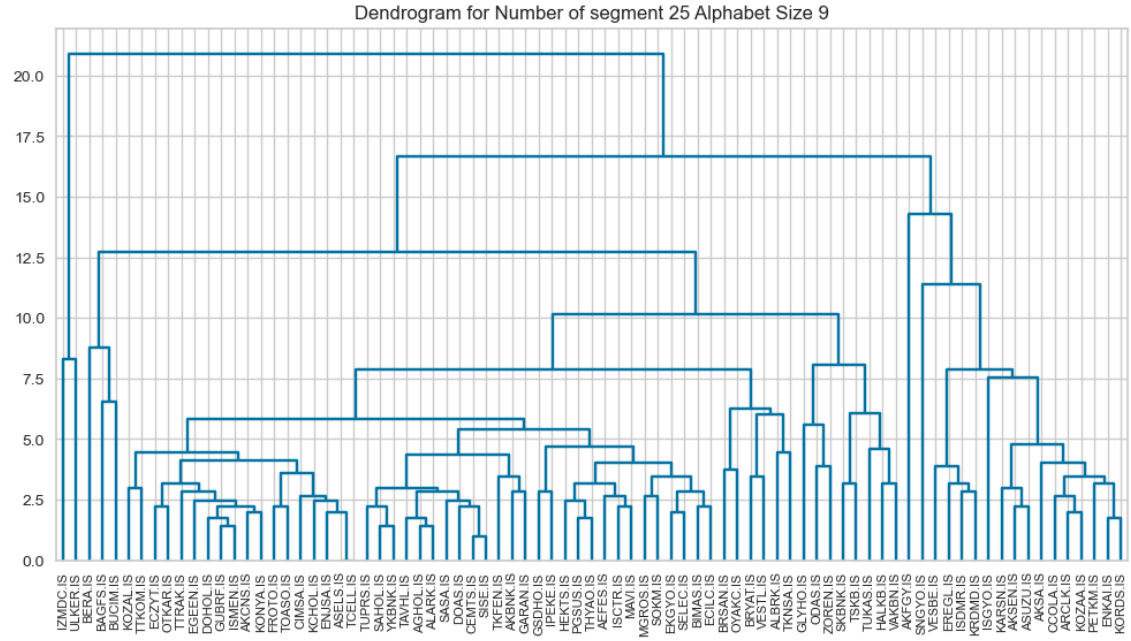
Appendix A 3 Dendrogram for Alphabet Size = 7 Number of Segment = 45 With Usage of Average Linkage



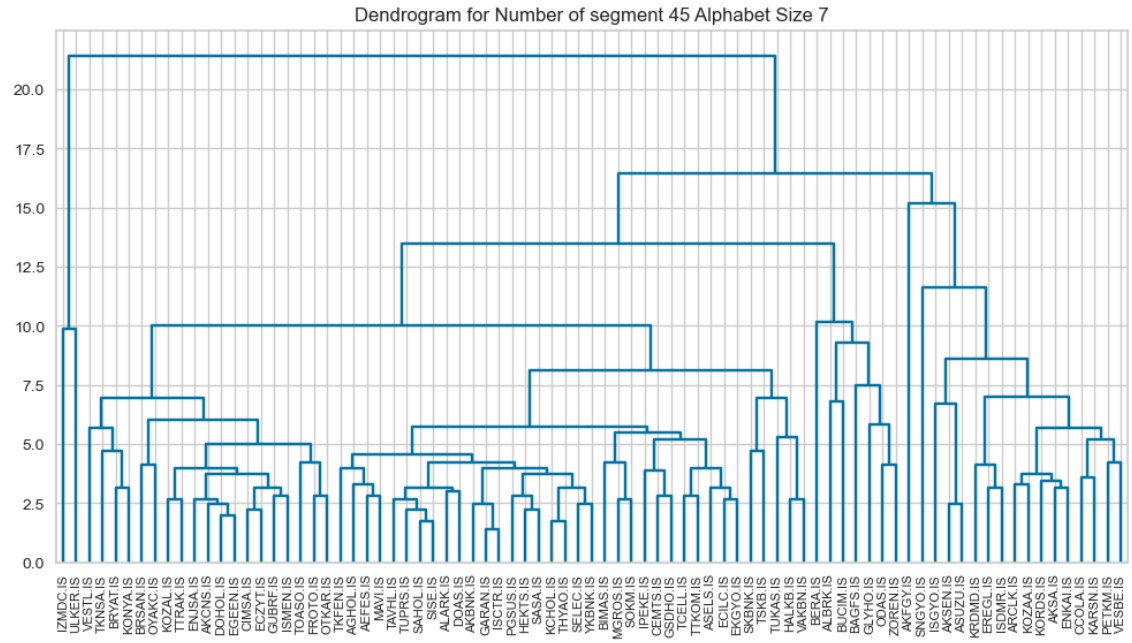
Appendix A 4 Dendrogram for Alphabet Size = 7 Number of Segment = 180 With Usage of Average Linkage



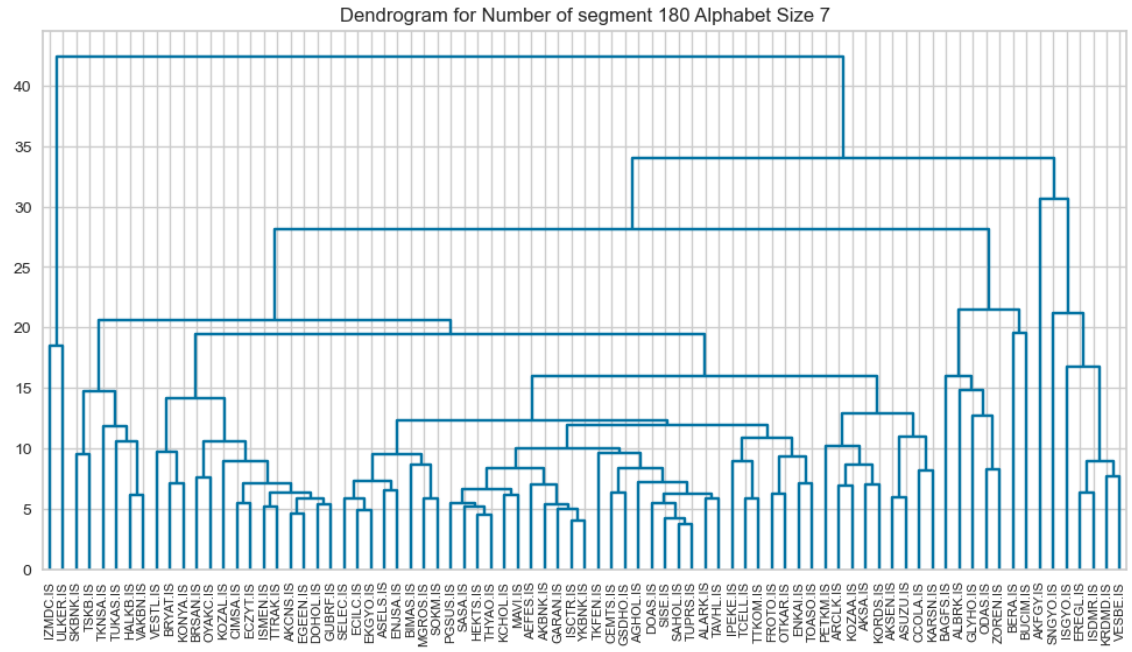
Appendix A 5 Dendrogram for Alphabet Size = 4 Number of Segment = 25 With Usage of Complete Linkage



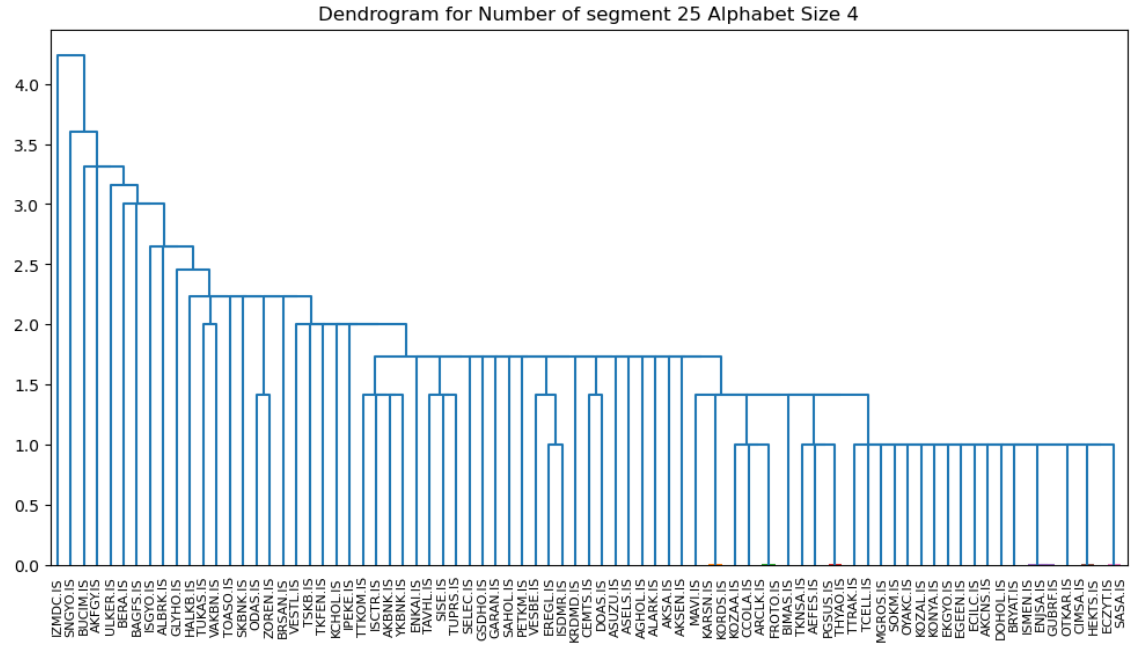
Appendix A 6 Dendrogram for Alphabet Size = 9 Number of Segment = 25 With Usage of Complete Linkage



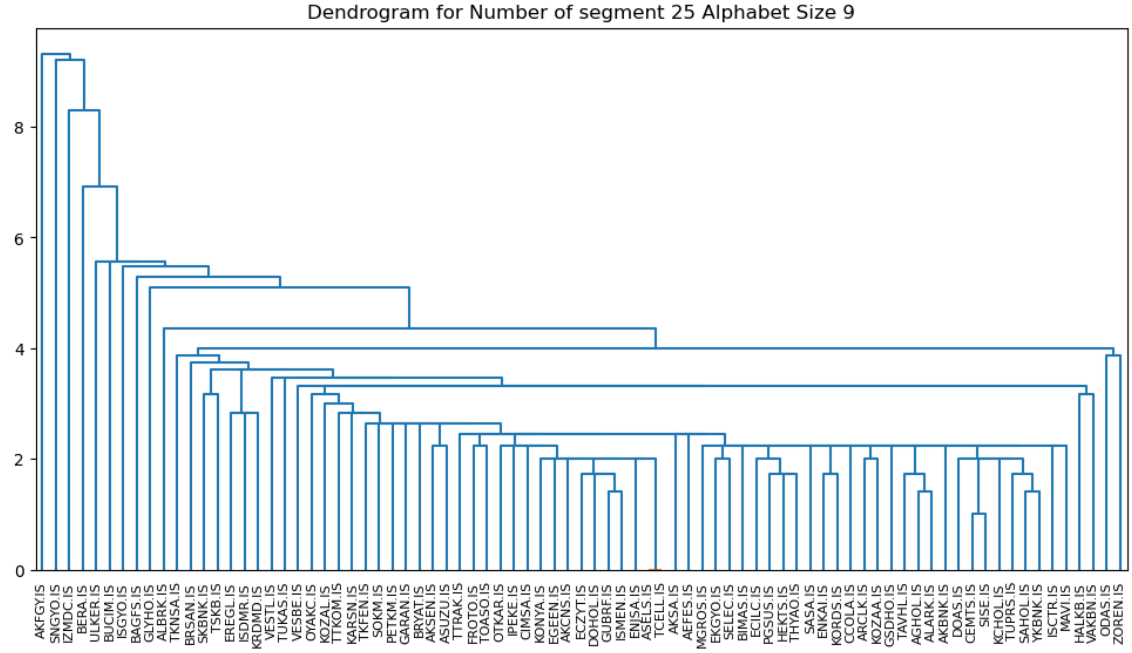
Appendix A 7 Dendrogram for Alphabet Size = 7 Number of Segment = 45 With Usage of Complete Linkage



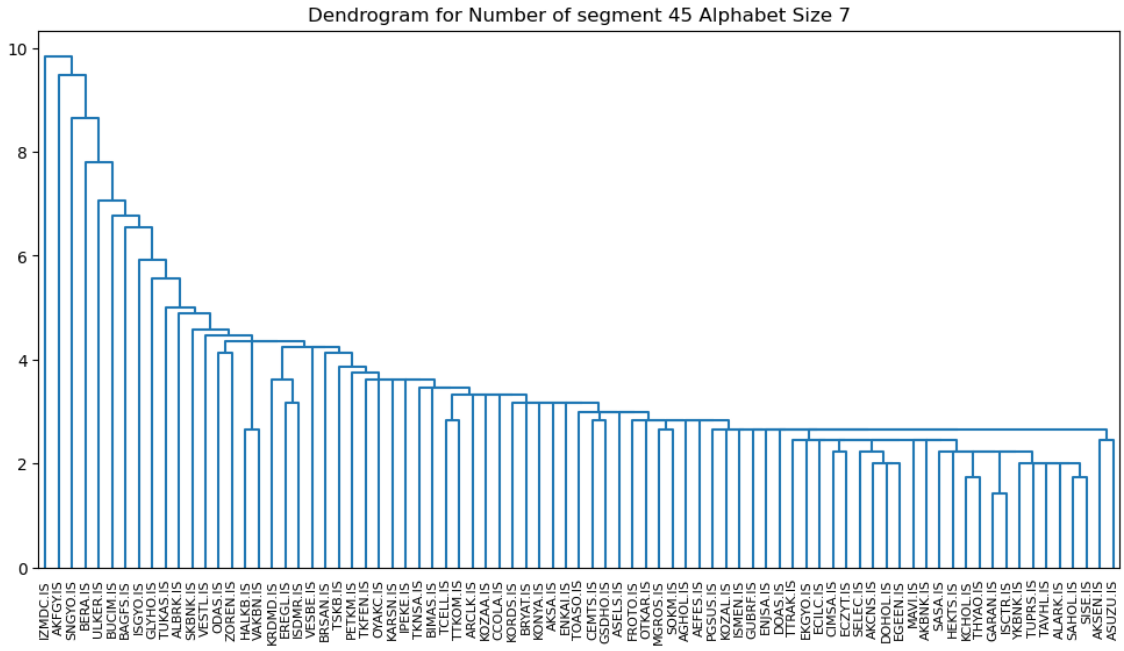
Appendix A 8 Dendrogram for Alphabet Size = 7 Number of Segment = 180 With Usage of Complete Linkage



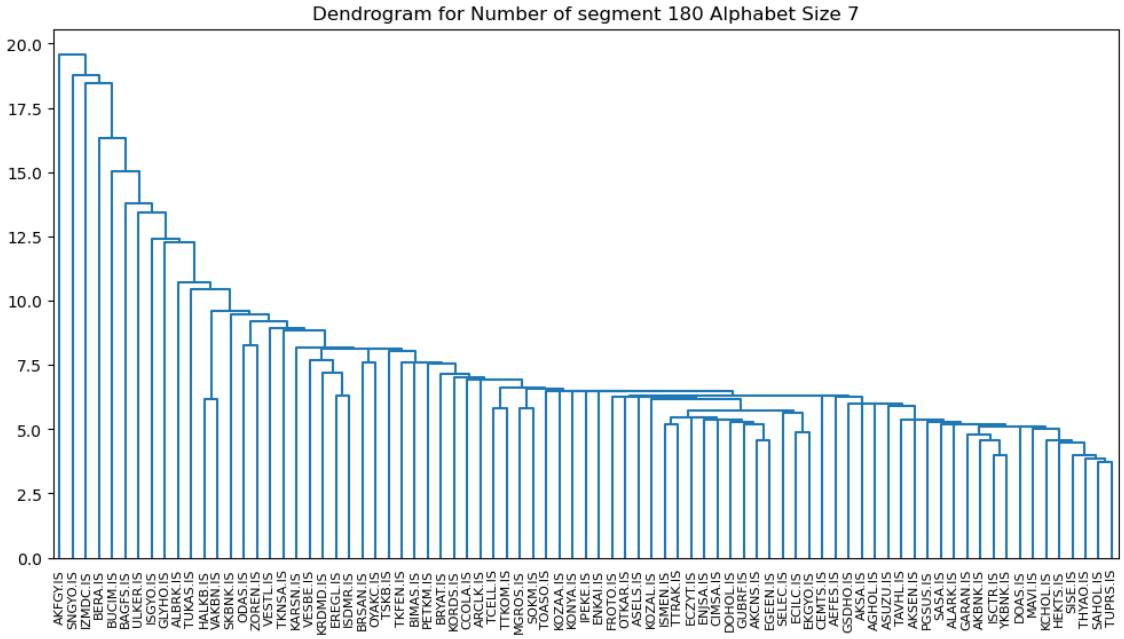
Appendix A 9 Dendrogram for Alphabet Size = 4 Number of Segment = 25 With Usage of Single Linkage



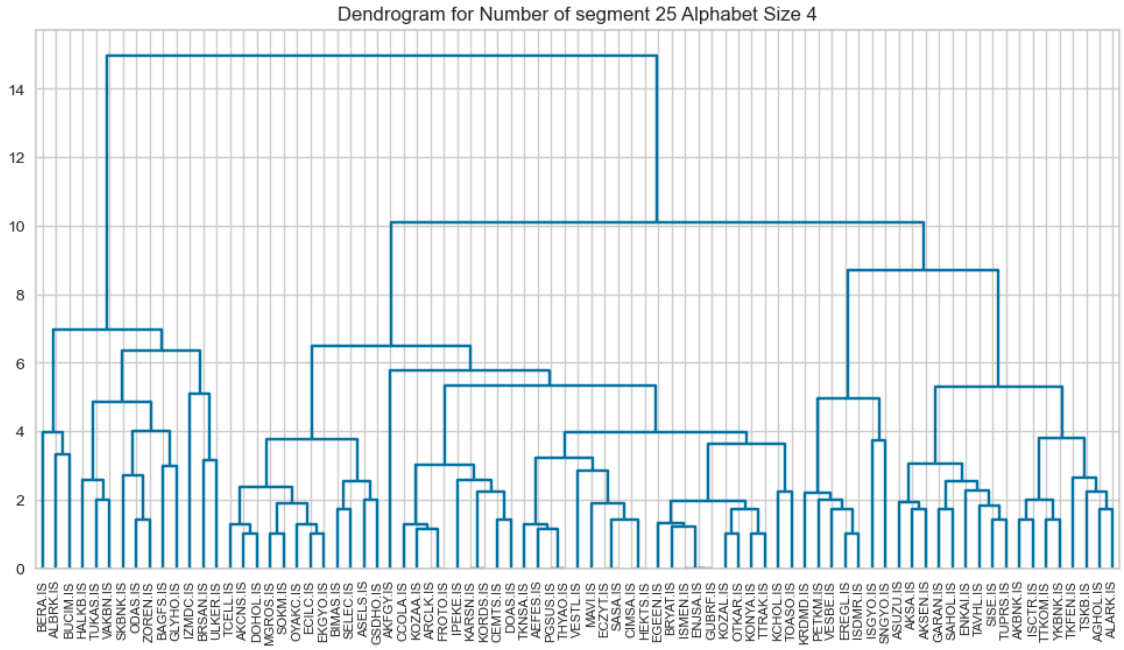
Appendix A 10 Dendrogram for Alphabet Size = 9 Number of Segment = 25 With Usage of Single Linkage



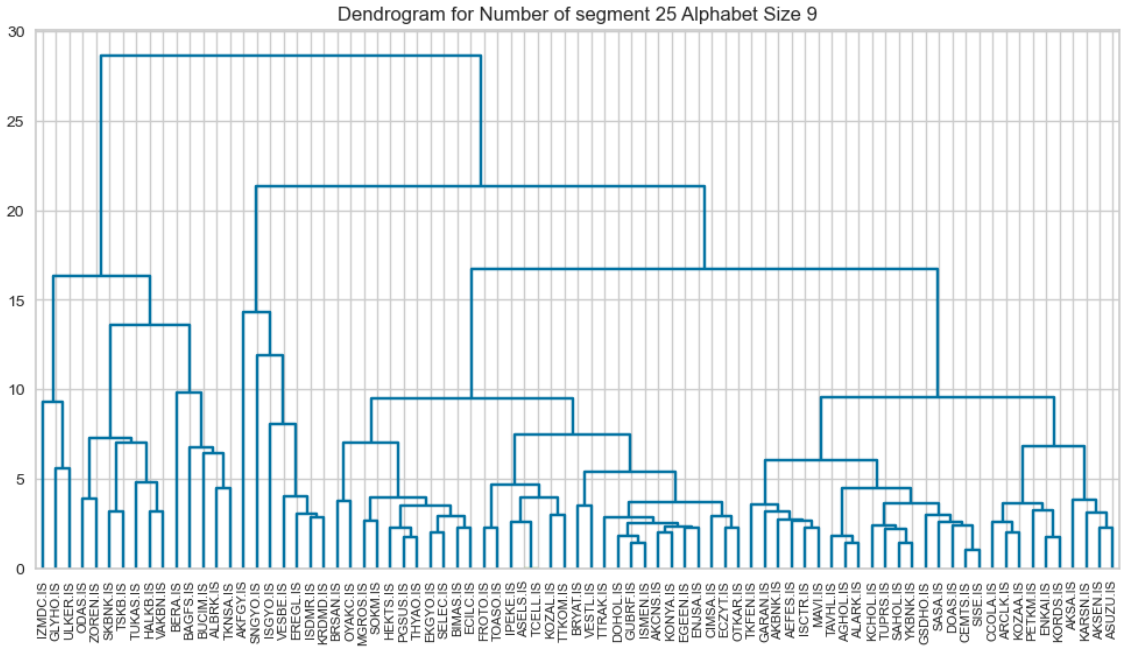
Appendix A 11 Dendrogram for Alphabet Size = 7 Number of Segment = 45 With Usage of Single Linkage



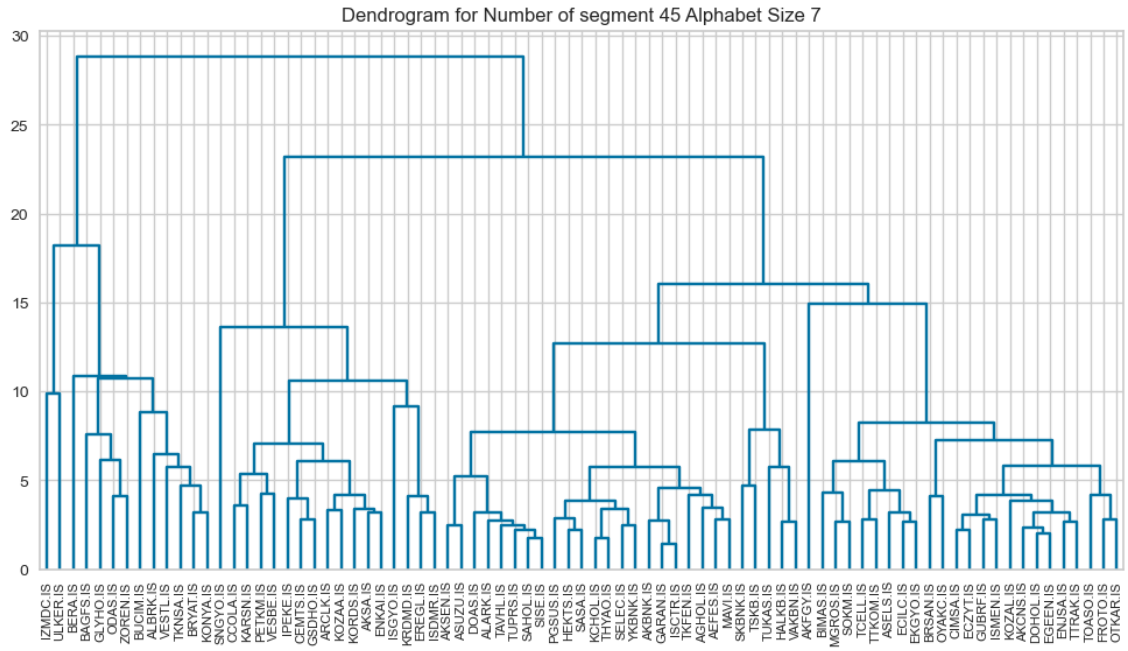
Appendix A 12 Dendrogram for Alphabet Size = 7 Number of Segment = 180 With Usage of Single Linkage



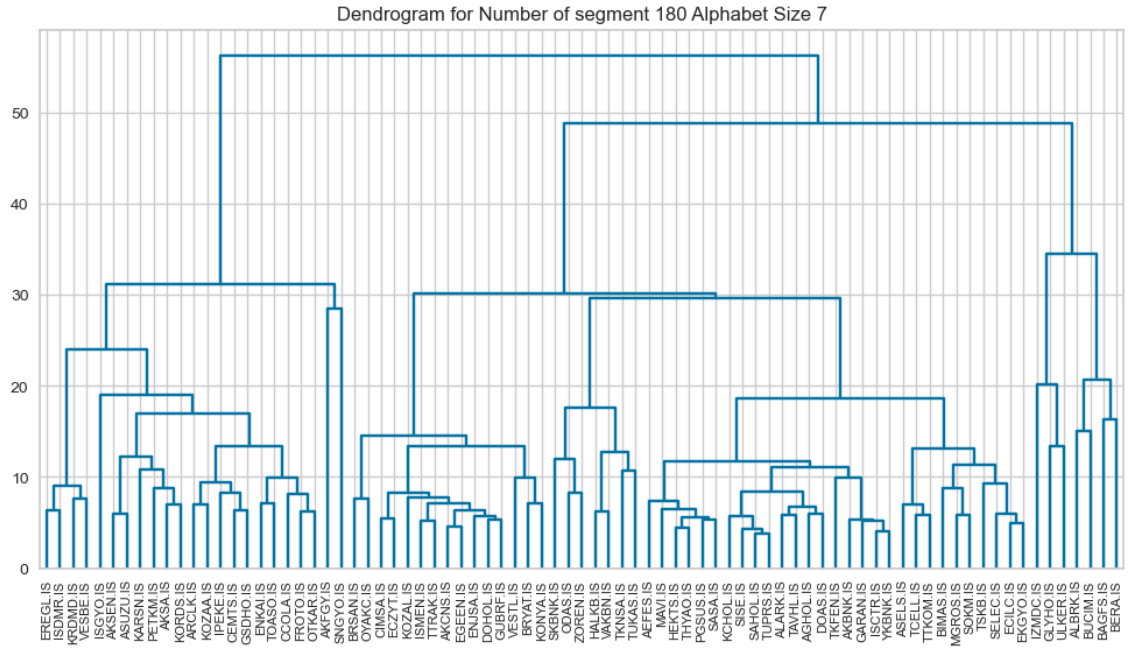
Appendix A 13 Dendrogram for Alphabet Size = 4 Number of Segment = 25 With Usage of Ward Linkage



Appendix A 14 Dendrogram for Alphabet Size = 9 Number of Segment = 25 With Usage of Ward Linkage



Appendix A 15 Dendrogram for Alphabet Size = 7 Number of Segment = 45 With Usage of Ward Linkage



Appendix A 16 Dendrogram for Alphabet Size = 7 Number of Segment = 180 With Usage of Ward Linkage

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS
AEFES	0	12	12	2	0	1	1	11	1	3	8	1	0	0	8	1	1	0	3	10	7	11
AGHOL	12	0	11	2	0	1	1	19	1	4	8	1	0	0	7	1	1	0	4	15	4	17
AKBNK	12	11	0	2	0	1	1	11	1	2	9	1	0	0	7	1	1	0	2	10	3	11
AKCNS	2	2	2	0	0	0	0	2	1	0	3	0	0	0	3	8	8	0	0	2	13	2
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	1	1	1	0	0	0	11	1	0	12	1	10	0	0	1	0	0	0	8	1	0	1
AKSEN	1	1	1	0	0	11	0	1	0	6	1	19	0	0	1	0	0	0	7	1	0	1
ALARK	11	19	11	2	0	1	1	0	1	4	8	1	0	0	7	1	1	0	4	15	4	18
ALBRK	1	1	1	1	0	0	0	1	0	0	1	0	2	0	1	2	2	4	0	1	1	1
ARCLK	3	4	2	0	0	12	6	4	0	0	2	6	0	0	1	0	0	0	14	5	2	5
ASELS	8	8	9	3	0	1	1	8	1	2	0	1	0	0	9	1	1	0	2	8	6	8
ASUZU	1	1	1	0	0	10	19	1	0	6	1	0	0	0	1	0	0	0	7	1	0	1
BAGFS	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	3	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	8	7	7	3	0	1	1	7	1	1	9	1	0	0	0	2	2	0	1	7	2	7
BRSAN	1	1	1	8	0	0	0	1	2	0	1	0	0	0	2	0	8	0	0	1	6	1
BRYAT	1	1	1	8	0	0	0	1	2	0	1	0	0	0	2	8	0	0	0	1	5	1
BUCIM	0	0	0	0	0	0	0	0	4	0	0	0	3	0	0	0	0	0	0	0	0	0
CCOLA	3	4	2	0	0	8	7	4	0	14	2	7	0	0	1	0	0	0	0	6	2	6
CEMTS	10	15	10	2	0	1	1	15	1	5	8	1	0	0	7	1	1	0	6	0	4	17
CIMSA	7	4	3	13	0	0	0	4	1	2	6	0	0	0	2	6	5	0	2	4	0	4
DOAS	11	17	11	2	0	1	1	18	1	5	8	1	0	0	7	1	1	0	6	17	4	0
DOHOL	2	2	2	20	0	0	0	2	1	0	3	0	0	0	3	8	8	0	0	2	13	2
ECILC	8	7	7	7	0	1	1	7	1	1	12	1	0	0	13	3	4	0	1	7	2	7
ECZYT	7	4	3	15	0	0	0	4	1	2	4	0	0	0	2	6	5	0	2	4	18	4
EGEEN	4	5	3	15	0	0	0	5	1	4	4	0	0	0	2	6	5	0	5	7	15	7
EKGYO	8	7	7	7	0	1	1	7	1	1	12	1	0	0	13	3	4	0	1	7	2	7
ENJSA	4	4	4	11	0	1	1	4	1	1	10	1	0	0	7	5	8	0	1	4	10	4
ENKAI	5	6	4	0	0	9	5	6	0	13	3	5	0	0	2	0	0	0	11	6	2	6
EREGL	0	0	0	0	0	3	3	0	0	2	0	3	0	0	0	0	0	0	2	0	0	0
FROTO	7	8	6	6	0	1	1	8	1	6	7	1	0	0	4	3	3	0	5	9	9	9
GARAN	11	10	15	2	0	4	4	10	1	1	7	4	0	0	7	1	1	0	1	9	2	10
GLYHO	0	0	0	0	0	0	0	0	3	0	0	0	5	0	0	0	0	1	0	0	0	0
GSDHO	10	10	10	2	0	1	1	10	1	2	11	1	0	0	8	1	1	0	2	13	3	10
GUBRF	2	2	2	18	0	0	0	2	1	0	3	0	0	0	3	8	10	0	0	2	13	2
HALKB	1	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1
HEKTS	17	11	11	2	0	1	1	11	1	3	8	1	0	0	8	1	1	0	3	10	7	11
IPEKE	11	12	9	2	0	1	1	12	1	4	8	1	0	0	8	1	1	0	4	13	4	11
ISCTR	15	10	17	2	0	1	1	10	1	2	9	1	0	0	8	1	1	0	2	9	3	10
ISDMR	0	0	0	0	0	3	3	0	0	2	0	3	0	0	0	0	0	0	2	0	0	0
ISGYO	0	0	0	0	0	5	6	0	0	3	0	6	0	0	0	0	0	0	3	0	0	0
ISMEN	2	2	2	18	0	0	0	2	1	0	3	0	0	0	3	8	10	0	0	2	13	2
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	3	4	2	0	0	8	9	4	0	12	2	9	0	0	1	0	0	0	18	6	2	6
KCHOL	13	11	11	3	0	1	1	11	1	3	11	1	0	0	7	1	1	0	3	10	8	11
KONYA	2	2	2	12	0	0	0	2	1	0	3	0	0	0	3	7	16	0	0	2	7	2
KORDS	3	4	2	0	0	13	6	4	0	17	2	6	0	0	1	0	0	0	14	6	2	6
KOZAA	3	4	2	0	0	12	6	4	0	20	2	6	0	0	1	0	0	0	14	5	2	5
KOZAL	2	2	2	15	0	0	0	2	1	0	3	0	0	0	3	8	10	0	0	2	12	2
KRDMD	0	0	0	0	0	3	3	0	0	2	0	3	0	0	0	0	0	0	2	0	0	0
MAVI	19	12	11	2	0	1	1	11	1	3	8	1	0	0	8	1	1	0	3	10	7	11
MGROS	8	7	7	6	0	1	1	7	1	1	9	1	0	0	13	3	4	0	1	7	2	7
ODAS	0	0	0	1	0	0	0	0	2	0	0	0	2	0	3	1	1	0	0	0	0	0
OTKAR	5	5	5	11	0	1	1	5	1	1	5	1	0	0	5	5	8	0	1	5	6	5
OYAKC	1	1	1	11	0	0	0	1	2	0	1	0	0	0	2	15	9	0	0	1	6	1
PETKM	1	1	1	0	0	10	7	1	0	10	1	7	0	0	1	0	0	0	9	1	0	1
PGSUS	17	11	11	2	0	1	1	11	1	3	8	1	0	0	8	1	1	0	3	10	7	11

Appendix B 1 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage

	DOHOL	ECILC	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN
AEFES	2	8	7	4	8	4	5	0	7	11	0	10	2	1	17	11	15	0	0	2	0	3
AGHOL	2	7	4	5	7	4	6	0	8	10	0	10	2	1	11	12	10	0	0	2	0	4
AKBNK	2	7	3	3	7	4	4	0	6	15	0	10	2	1	11	9	17	0	0	2	0	2
AKCNS	20	7	15	15	7	11	0	0	6	2	0	2	18	0	2	2	2	0	0	18	0	0
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	1	0	0	1	1	9	3	1	4	0	1	0	0	1	1	1	3	5	0	0	8
AKSEN	0	1	0	0	1	1	5	3	1	4	0	1	0	0	1	1	1	3	6	0	0	9
ALARK	2	7	4	5	7	4	6	0	8	10	0	10	2	1	11	12	10	0	0	2	0	4
ALBRK	1	1	1	1	1	1	0	0	1	1	3	1	1	0	1	1	1	0	0	1	0	0
ARCLK	0	1	2	4	1	1	13	2	6	1	0	2	0	0	3	4	2	2	3	0	0	12
ASELS	3	12	4	4	12	10	3	0	7	7	0	11	3	1	8	8	9	0	0	3	0	2
ASUZU	0	1	0	0	1	1	5	3	1	4	0	1	0	0	1	1	1	3	6	0	0	9
BAGFS	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	3	13	2	2	13	7	2	0	4	7	0	8	3	1	8	8	8	0	0	3	0	1
BRSAN	8	3	6	6	3	5	0	0	3	1	0	1	8	0	1	1	1	0	0	8	0	0
BRYAT	8	4	5	5	4	8	0	0	3	1	0	1	10	0	1	1	1	0	0	10	0	0
BUCIM	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CCOLA	0	1	2	5	1	1	11	2	5	1	0	2	0	0	3	4	2	2	3	0	0	18
CEMTS	2	7	4	7	7	4	6	0	9	9	0	13	2	1	10	13	9	0	0	2	0	6
CIMSA	13	2	18	15	2	10	2	0	9	2	0	3	13	0	7	4	3	0	0	13	0	2
DOAS	2	7	4	7	7	4	6	0	9	10	0	10	2	1	11	11	10	0	0	2	0	6
DOHOL	0	7	15	15	7	11	0	0	6	2	0	2	18	0	2	2	2	0	0	18	0	0
ECILC	7	0	2	2	20	10	2	0	4	7	0	8	5	1	8	8	8	0	0	5	0	1
ECZYT	15	2	0	17	2	8	2	0	8	2	0	3	15	0	7	4	3	0	0	15	0	2
EGEEN	15	2	17	0	2	8	3	0	10	2	0	3	15	0	4	5	3	0	0	15	0	5
EKGYO	7	20	2	2	0	10	2	0	4	7	0	8	5	1	8	8	8	0	0	5	0	1
ENJSA	11	10	8	8	10	0	2	0	9	4	0	4	13	0	4	4	4	0	0	13	0	1
ENKAI	0	2	2	3	2	2	0	2	7	3	0	4	0	0	5	6	4	2	3	0	0	10
EREGL	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	20	4	0	0	2
FROTO	6	4	8	10	4	9	7	0	0	5	0	6	6	0	7	8	6	0	0	6	0	5
GARAN	2	7	2	2	7	4	3	1	5	0	0	8	2	1	10	8	12	1	2	2	0	1
GLYHO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GSDHO	2	8	3	3	8	4	4	0	6	8	0	0	2	1	10	14	11	0	0	2	0	2
GUBRF	18	5	15	15	5	13	0	0	6	2	0	2	0	0	2	2	2	0	0	20	0	0
HALKB	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1	1	1	0	0	0	0	0
HEKTS	2	8	7	4	8	4	5	0	7	10	0	10	2	1	0	11	13	0	0	2	0	3
IPEKE	2	8	4	5	8	4	6	0	8	8	0	14	2	1	11	0	10	0	0	2	0	4
ISCTR	2	8	3	3	8	4	4	0	6	12	0	11	2	1	13	10	0	0	0	2	0	2
ISDMR	0	0	0	0	0	0	2	20	0	1	0	0	0	0	0	0	0	0	4	0	0	2
ISGYO	0	0	0	0	0	0	3	4	0	2	0	0	0	0	0	0	0	4	0	0	0	3
ISMEN	18	5	15	15	5	13	0	0	6	2	0	2	20	0	2	2	2	0	0	0	0	0
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	0	1	2	5	1	1	10	2	5	1	0	2	0	0	3	4	2	2	3	0	0	0
KCHOL	3	7	6	5	7	7	5	0	9	10	0	9	3	1	15	10	11	0	0	3	0	3
KONYA	12	5	9	9	5	10	0	0	5	2	0	2	14	0	2	2	2	0	0	14	0	0
KORDS	0	1	2	5	1	1	14	2	5	1	0	2	0	0	3	4	2	2	3	0	0	13
KOZAA	0	1	2	4	1	1	13	2	6	1	0	2	0	0	3	4	2	2	3	0	0	12
KOZAL	15	5	12	12	5	13	0	0	6	2	0	2	17	0	2	2	2	0	0	17	0	0
KRDMD	0	0	0	0	0	0	2	16	0	1	0	0	0	0	0	0	0	16	4	0	0	2
MAVI	2	8	7	4	8	4	5	0	7	10	0	10	2	1	18	11	14	0	0	2	0	3
MGROS	6	15	2	2	15	9	2	0	4	7	0	8	5	1	8	8	8	0	0	5	0	1
ODAS	1	1	0	0	1	1	0	0	0	0	6	0	1	0	0	0	0	0	0	1	0	0
OTKAR	11	7	8	8	7	13	4	0	13	5	0	5	13	0	5	5	5	0	0	13	0	1
OYAKC	11	6	6	6	6	6	0	0	3	1	0	1	9	0	1	1	1	0	0	9	0	0
PETKM	0	1	0	0	1	1	9	7	1	2	0	1	0	0	1	1	1	7	4	0	0	8
PGSUS	2	8	7	4	8	4	5	0	7	10	0	10	2	1	20	11	13	0	0	2	0	3

Appendix B 2 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	KCHOL	KONYA	KORDS	KOZAA	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL
AEFES	13	2	3	3	2	0	19	8	0	5	1	1	17	9	15	9	9	1	0	8	9	8
AGHOL	11	2	4	4	2	0	12	7	0	5	1	1	11	14	14	8	14	1	0	7	14	8
AKBNK	11	2	2	2	2	0	11	7	0	5	1	1	11	10	12	8	10	1	0	7	10	8
AKCNS	3	12	0	0	15	0	2	6	1	11	11	0	2	2	2	3	2	1	0	6	2	8
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	1	0	13	12	0	3	1	1	0	1	0	10	1	4	1	1	4	0	0	1	5	1
AKSEN	1	0	6	6	0	3	1	1	0	1	0	7	1	4	1	1	4	0	0	1	5	1
ALARK	11	2	4	4	2	0	11	7	0	5	1	1	11	15	14	8	15	1	0	7	15	8
ALBRK	1	1	0	0	1	0	1	1	2	1	2	0	1	1	1	1	1	0	0	1	1	1
ARCLK	3	0	17	20	0	2	3	1	0	1	0	10	3	1	3	1	1	0	0	1	1	1
ASELS	11	3	2	2	3	0	8	9	0	5	1	1	8	7	8	10	7	1	0	9	7	12
ASUZU	1	0	6	6	0	3	1	1	0	1	0	7	1	4	1	1	4	0	0	1	6	1
BAGFS	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	7	3	1	1	3	0	8	13	3	5	2	1	8	7	7	17	7	4	0	13	7	8
BRSAN	1	7	0	0	8	0	1	3	1	5	15	0	1	1	1	2	1	1	0	3	1	3
BRYAT	1	16	0	0	10	0	1	4	1	8	9	0	1	1	1	2	1	1	0	4	1	4
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	3	0	14	14	0	2	3	1	0	1	0	9	3	1	3	1	1	0	0	1	1	1
CEMTS	10	2	6	5	2	0	10	7	0	5	1	1	10	12	13	7	12	1	0	7	12	8
CIMSA	8	7	2	2	12	0	7	2	0	6	6	0	7	2	7	2	2	0	0	2	2	5
DOAS	11	2	6	5	2	0	11	7	0	5	1	1	11	15	14	8	15	1	0	7	15	8
DOHOL	3	12	0	0	15	0	2	6	1	11	11	0	2	2	2	3	2	1	0	6	2	8
ECILC	7	5	1	1	5	0	8	15	1	7	6	1	8	7	7	14	7	2	0	15	7	14
ECZYT	6	9	2	2	12	0	7	2	0	8	6	0	7	2	7	2	2	0	0	2	2	3
EGEEN	5	9	5	4	12	0	4	2	0	8	6	0	4	2	4	2	2	0	0	2	2	3
EKGYO	7	5	1	1	5	0	8	15	1	7	6	1	8	7	7	14	7	2	0	15	7	14
ENJSA	7	10	1	1	13	0	4	9	1	13	6	1	4	4	4	8	4	1	0	9	4	10
ENKAI	5	0	14	13	0	2	5	2	0	4	0	9	5	3	5	2	3	0	0	2	3	3
EREGL	0	0	2	2	0	16	0	0	0	0	0	7	0	1	0	0	1	0	0	0	1	0
FROTO	9	5	5	6	6	0	7	4	0	13	3	1	7	5	7	4	5	0	0	4	5	7
GARAN	10	2	1	1	2	1	10	7	0	5	1	2	10	15	11	8	15	1	0	7	13	8
GLYHO	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
GSDHO	9	2	2	2	2	0	10	8	0	5	1	1	10	9	9	8	9	1	0	8	9	8
GUBRF	3	14	0	0	17	0	2	5	1	13	9	0	2	2	2	3	2	1	0	5	2	6
HALKB	1	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	6	0	1	1	1
HEKTS	15	2	3	3	2	0	18	8	0	5	1	1	20	9	17	10	9	1	0	8	9	8
IPEKE	10	2	4	4	2	0	11	8	0	5	1	1	11	8	10	8	8	1	0	8	8	9
ISCTR	11	2	2	2	2	0	14	8	0	5	1	1	13	9	11	9	9	1	0	8	9	8
ISDMR	0	0	2	2	0	16	0	0	0	0	0	7	0	1	0	0	1	0	0	0	1	0
ISGYO	0	0	3	3	0	4	0	0	0	0	0	4	0	2	0	0	2	0	0	0	2	0
ISMEN	3	14	0	0	17	0	2	5	1	13	9	0	2	2	2	3	2	1	0	5	2	6
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	3	0	13	12	0	2	3	1	0	1	0	8	3	1	3	1	1	0	0	1	1	1
KCHOL	0	3	3	3	3	0	14	7	0	6	1	1	15	9	15	9	9	1	0	7	9	11
KONYA	3	0	0	0	12	0	2	5	1	12	8	0	2	2	2	3	2	1	0	5	2	6
KORDS	3	0	0	17	0	2	3	1	0	1	0	11	3	1	3	1	1	0	0	1	1	1
KOZAA	3	0	17	0	0	2	3	1	0	1	0	10	3	1	3	1	1	0	0	1	1	1
KOZAL	3	12	0	0	0	0	2	5	1	11	9	0	2	2	2	3	2	1	0	5	2	6
KRDMD	0	0	2	2	0	0	0	0	0	0	0	5	0	1	0	0	1	0	0	0	1	0
MAVI	14	2	3	3	2	0	0	8	0	5	1	1	18	9	16	9	9	1	0	8	9	8
MGROS	7	5	1	1	5	0	8	0	1	7	5	1	8	7	7	12	7	2	0	20	7	11
ODAS	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	3	0	3	0	1	0	1
OTKAR	6	12	1	1	11	0	5	7	1	0	6	1	5	5	5	5	5	1	0	7	5	9
OYAKC	1	8	0	0	9	0	1	5	1	6	0	0	1	1	1	2	1	1	0	5	1	6
PETKM	1	0	11	10	0	5	1	1	0	1	0	0	1	2	1	1	2	0	0	1	2	1
PGSUS	15	2	3	3	2	0	18	8	0	5	1	1	0	9	17	10	9	1	0	8	9	8

Appendix B 3 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
AEFES	17	10	6	8	2	9	2	1	9	0	1	0	1	12	0
AGHOL	11	11	3	7	2	9	2	1	14	0	1	0	1	13	0
AKBNK	11	15	2	6	3	12	2	1	10	0	1	0	1	16	0
AKCNS	2	2	3	6	0	3	18	0	2	0	0	0	5	2	1
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	1	1	0	1	0	1	0	0	4	0	0	3	0	1	0
AKSEN	1	1	0	1	0	1	0	0	4	0	0	3	0	1	0
ALARK	11	11	3	7	2	9	2	1	15	0	1	0	1	13	0
ALBRK	1	1	3	1	0	1	1	0	1	0	0	0	2	1	2
ARCLK	3	2	2	3	1	2	0	0	1	0	0	4	0	2	0
ASELS	8	9	2	7	5	12	3	1	7	0	1	0	1	9	0
ASUZU	1	1	0	1	0	1	0	0	4	0	0	3	0	1	0
BAGFS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
BERA	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
BIMAS	8	7	1	4	1	7	3	1	7	0	1	0	1	7	3
BRSAN	1	1	4	3	0	1	8	0	1	0	0	0	6	1	1
BRYAT	1	1	5	3	0	1	10	0	1	0	0	0	10	1	1
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	3	2	2	3	1	2	0	0	1	0	0	4	0	2	0
CEMTS	10	10	3	7	2	9	2	1	12	0	1	0	1	12	0
CIMSA	7	3	8	10	1	4	13	0	2	0	0	0	5	3	0
DOAS	11	11	3	7	2	9	2	1	15	0	1	0	1	13	0
DOHOL	2	2	3	6	0	3	18	0	2	0	0	0	5	2	1
ECILC	8	7	1	4	1	9	5	1	7	0	1	0	1	7	1
ECZYT	7	3	8	9	1	4	15	0	2	0	0	0	5	3	0
EGEEN	4	3	5	8	1	4	15	0	2	0	0	0	5	3	0
EKGYO	8	7	1	4	1	9	5	1	7	0	1	0	1	7	1
ENJSA	4	4	3	9	0	5	13	0	4	0	0	0	3	4	1
ENKAI	5	4	2	7	1	4	0	0	3	0	0	4	0	4	0
EREGL	0	0	0	0	0	0	0	0	1	0	0	12	0	0	0
FROTO	7	6	5	15	1	7	6	0	5	0	0	0	3	6	0
GARAN	10	11	1	5	1	8	2	1	15	0	1	1	1	12	0
GLYHO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
GSDHO	10	10	2	6	5	10	2	1	9	0	1	0	1	10	0
GUBRF	2	2	3	6	0	3	20	0	2	0	0	0	5	2	1
HALKB	1	1	3	0	6	1	0	12	1	0	17	0	0	1	0
HEKTS	20	10	6	8	2	9	2	1	9	0	1	0	1	12	0
IPEKE	11	9	3	7	2	10	2	1	8	0	1	0	1	9	0
ISCTR	13	13	2	6	3	12	2	1	9	0	1	0	1	15	0
ISDMR	0	0	0	0	0	0	0	0	1	0	0	12	0	0	0
ISGYO	0	0	0	0	0	0	0	0	2	0	0	4	0	0	0
ISMEN	2	2	3	6	0	3	20	0	2	0	0	0	5	2	1
IZMDC	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
KARSN	3	2	2	3	1	2	0	0	1	0	0	4	0	2	0
KCHOL	15	10	4	11	2	10	3	1	9	0	1	0	1	12	0
KONYA	2	2	4	5	0	3	14	0	2	0	0	0	7	2	1
KORDS	3	2	2	3	1	2	0	0	1	0	0	4	0	2	0
KOZAA	3	2	2	3	1	2	0	0	1	0	0	4	0	2	0
KOZAL	2	2	3	6	0	5	17	0	2	0	0	0	5	2	1
KRDMD	0	0	0	0	0	0	0	0	1	0	0	15	0	0	0
MAVI	18	10	6	8	2	9	2	1	9	0	1	0	1	11	0
MGROS	8	7	1	4	1	7	5	1	7	0	1	0	1	7	1
ODAS	0	0	0	0	0	0	1	0	0	0	0	0	0	0	17
OTKAR	5	5	3	11	0	6	13	0	5	0	0	0	3	5	1
OYAKC	1	1	4	3	0	1	9	0	1	0	0	0	6	1	1
PETKM	1	1	0	1	0	1	0	0	2	0	0	8	0	1	0
PGSUS	20	10	6	8	2	9	2	1	9	0	1	0	1	12	0

Appendix B 4 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	COOJA	CEMTS	CIMSA	DOAS
SAHOL	9	14	10	2	0	4	4	15	1	1	7	4	0	0	7	1	1	0	1	12	2	15
SASA	15	14	12	2	0	1	1	14	1	3	8	1	0	0	7	1	1	0	3	13	7	14
SELEC	9	8	8	3	0	1	1	8	1	1	10	1	0	0	17	2	2	0	1	7	2	8
SISE	9	14	10	2	0	4	4	15	1	1	7	4	0	0	7	1	1	0	1	12	2	15
SKBNK	1	1	1	1	0	0	0	1	0	0	1	0	0	0	4	1	1	0	0	1	0	1
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	8	7	7	6	0	1	1	7	1	1	9	1	0	0	13	3	4	0	1	7	2	7
TAVHL	9	14	10	2	0	5	5	15	1	1	7	6	0	0	7	1	1	0	1	12	2	15
TCELL	8	8	8	8	0	1	1	8	1	1	12	1	0	0	8	3	4	0	1	8	5	8
THYAO	17	11	11	2	0	1	1	11	1	3	8	1	0	0	8	1	1	0	3	10	7	11
TKFEN	10	11	15	2	0	1	1	11	1	2	9	1	0	0	7	1	1	0	2	10	3	11
TKNSA	6	3	2	3	0	0	0	3	3	2	2	0	0	0	1	4	5	0	2	3	8	3
TOASO	8	7	6	6	0	1	1	7	1	3	7	1	0	0	4	3	3	0	3	7	10	7
TSKB	2	2	3	0	0	0	0	2	0	1	5	0	0	0	1	0	0	0	1	2	1	2
TTKOM	9	9	12	3	0	1	1	9	1	2	12	1	0	0	7	1	1	0	2	9	4	9
TTRAK	2	2	2	18	0	0	0	2	1	0	3	0	0	0	3	8	10	0	0	2	13	2
TUKAS	1	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1
TUPRS	9	14	10	2	0	4	4	15	1	1	7	4	0	0	7	1	1	0	1	12	2	15
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	1	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	1
VESBE	0	0	0	0	0	5	3	0	0	4	0	3	0	0	0	0	0	0	4	0	0	0
VESTL	1	1	1	5	0	0	0	1	2	0	1	0	0	2	1	6	10	0	0	1	5	1
YKBNK	12	13	16	2	0	1	1	13	1	2	9	1	0	0	7	1	1	0	2	12	3	13
ZOREN	0	0	0	1	0	0	0	0	2	0	0	0	2	0	3	1	1	0	0	0	0	0

Appendix B 5 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	DOHOL	EGLIC	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN
SAHOL	2	7	2	2	7	4	3	1	5	15	0	9	2	1	9	8	9	1	2	2	0	1
SASA	2	7	7	4	7	4	5	0	7	11	0	9	2	1	17	10	11	0	0	2	0	3
SELEC	3	14	2	2	14	8	2	0	4	8	0	8	3	1	10	8	9	0	0	3	0	1
SISE	2	7	2	2	7	4	3	1	5	15	0	9	2	1	9	8	9	1	2	2	0	1
SKBNK	1	2	0	0	2	1	0	0	0	1	0	1	1	6	1	1	1	0	0	1	0	0
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	6	15	2	2	15	9	2	0	4	7	0	8	5	1	8	8	8	0	0	5	0	1
TAVHL	2	7	2	2	7	4	3	1	5	13	0	9	2	1	9	8	9	1	2	2	0	1
TCELL	8	14	3	3	14	10	3	0	7	8	0	8	6	1	8	9	8	0	0	6	0	1
THYAO	2	8	7	4	8	4	5	0	7	10	0	10	2	1	20	11	13	0	0	2	0	3
TKFEN	2	7	3	3	7	4	4	0	6	11	0	10	2	1	10	9	13	0	0	2	0	2
TKNSA	3	1	8	5	1	3	2	0	5	1	0	2	3	3	6	3	2	0	0	3	0	2
TOASO	6	4	9	8	4	9	7	0	15	5	0	6	6	0	8	7	6	0	0	6	0	3
TSKB	0	1	1	1	1	0	1	0	1	1	0	5	0	6	2	2	3	0	0	0	0	1
TTKOM	3	9	4	4	9	5	4	0	7	8	0	10	3	1	9	10	12	0	0	3	0	2
TTRAK	18	5	15	15	5	13	0	0	6	2	0	2	20	0	2	2	2	0	0	20	0	0
TUKAS	0	1	0	0	1	0	0	0	0	1	0	1	0	12	1	1	1	0	0	0	0	0
TUPRS	2	7	2	2	7	4	3	1	5	15	0	9	2	1	9	8	9	1	2	2	0	1
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
VAKBN	0	1	0	0	1	0	0	0	0	1	0	1	0	17	1	1	1	0	0	0	0	0
VESBE	0	0	0	0	0	0	4	12	0	1	0	0	0	0	0	0	0	12	4	0	0	4
VESTL	5	1	5	5	1	3	0	0	3	1	0	1	5	0	1	1	1	0	0	5	0	0
YKBNK	2	7	3	3	7	4	4	0	6	12	0	10	2	1	12	9	15	0	0	2	0	2
ZOREN	1	1	0	0	1	1	0	0	0	0	6	0	1	0	0	0	0	0	0	1	0	0

Appendix B 6 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	KCHOL	KONYA	KORDS	KOZAA	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
SAHOL	9	2	1	1	2	1	9	7	0	5	1	2	9	0	12	8	20	1	0	7	18	8	9	10	1	5	1	8	2	1	20	0	1	1	1	12	0
SASA	15	2	3	3	2	0	16	7	0	5	1	1	17	12	0	9	12	1	0	7	12	8	17	11	6	8	2	9	2	1	12	0	1	0	1	15	0
SELEC	9	3	1	1	3	0	9	12	3	5	2	1	10	8	9	0	8	4	0	12	8	8	10	8	1	4	1	7	3	1	8	0	1	0	1	9	3
SISE	9	2	1	1	2	1	9	7	0	5	1	2	9	20	12	8	0	1	0	7	18	8	9	10	1	5	1	8	2	1	20	0	1	1	1	12	0
SKBNK	1	1	0	0	1	0	1	2	3	1	1	0	1	1	1	4	1	0	0	2	1	2	1	1	2	0	11	1	1	6	1	0	6	0	0	1	3
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	7	5	1	1	5	0	8	20	1	7	5	1	8	7	7	12	7	2	0	0	7	11	8	7	1	4	1	7	5	1	7	0	1	0	1	7	1
TAVHL	9	2	1	1	2	1	9	7	0	5	1	2	9	18	12	8	18	1	0	7	0	8	9	10	1	5	1	8	2	1	18	0	1	1	1	12	0
TCELL	11	6	1	1	6	0	8	11	1	9	6	1	8	8	8	8	8	2	0	11	8	0	8	8	1	7	1	13	6	1	8	0	1	0	1	8	1
THYAO	15	2	3	3	2	0	18	8	0	5	1	1	20	9	17	10	9	1	0	8	9	8	0	10	6	8	2	9	2	1	9	0	1	0	1	12	0
TKFEN	10	2	2	2	2	0	10	7	0	5	1	1	10	10	11	8	10	1	0	7	10	8	10	0	2	6	3	12	2	1	10	0	1	0	1	14	0
TKNSA	4	4	2	2	3	0	6	1	0	3	4	0	6	1	6	1	1	2	0	1	1	1	6	2	0	6	3	2	3	3	1	0	3	0	5	2	0
TOASO	11	5	3	3	6	0	8	4	0	11	3	1	8	5	8	4	5	0	0	4	5	7	8	6	6	0	1	7	6	0	5	0	0	0	3	6	0
TSKB	2	0	1	1	0	0	2	1	0	0	0	0	2	1	2	1	1	11	0	1	1	1	2	3	3	1	0	3	0	6	1	0	6	0	0	3	0
TTKOM	10	3	2	2	5	0	9	7	0	6	1	1	9	8	9	7	8	1	0	7	8	13	9	12	2	7	3	0	3	1	8	0	1	0	1	13	0
TTRAK	3	14	0	0	17	0	2	5	1	13	9	0	2	2	2	3	2	1	0	5	2	6	2	2	3	6	0	3	0	0	2	0	0	0	5	2	1
TUKAS	1	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	6	0	1	1	1	1	1	3	0	6	1	0	0	1	0	13	0	0	1	0
TUPRS	9	2	1	1	2	1	9	7	0	5	1	2	9	20	12	8	20	1	0	7	18	8	9	10	1	5	1	8	2	1	0	0	1	1	1	12	0
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	1	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	6	0	1	1	1	1	1	3	0	6	1	0	13	1	0	0	0	0	1	0
VESBE	0	0	4	4	0	15	0	0	0	0	0	8	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
VESTL	1	7	0	0	5	0	1	1	0	3	6	0	1	1	1	1	1	0	0	1	1	1	1	1	5	3	0	1	5	0	1	0	0	0	0	1	0
YKBNK	12	2	2	2	2	0	11	7	0	5	1	1	12	12	15	9	12	1	0	7	12	8	12	14	2	6	3	13	2	1	12	0	1	0	1	0	0
ZOREN	0	1	0	0	1	0	0	1	17	1	1	0	0	0	0	3	0	3	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Appendix B 7 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Complete Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL	ECILC
AEFES	0	14	17	13	0	8	10	14	1	9	12	7	0	0	11	5	11	0	9	11	12	15	13	14
AGHOL	14	0	14	10	0	9	11	20	1	10	12	8	0	0	11	5	9	0	10	13	10	17	10	11
AKBNK	17	14	0	11	0	8	10	14	1	8	12	7	0	0	11	5	9	0	8	10	11	14	11	12
AKCNS	13	10	11	0	0	8	8	10	1	9	14	7	0	0	11	5	12	0	9	11	18	11	20	15
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	8	9	8	8	0	0	14	9	1	13	8	12	0	0	8	5	7	0	11	9	8	9	8	8
AKSEN	10	11	10	8	0	14	0	11	1	11	8	16	0	0	8	5	7	0	11	9	8	11	8	8
ALARK	14	20	14	10	0	9	11	0	1	10	12	8	0	0	11	5	9	0	10	13	10	17	10	11
ALBRK	1	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	2	4	1	1	1	1	1	1
ARCLK	9	10	8	9	0	13	11	10	1	0	8	10	0	0	8	5	8	0	16	12	10	12	9	9
ASELS	12	12	12	14	0	8	8	12	1	8	0	7	0	0	12	5	9	0	8	10	14	11	14	15
ASUZU	7	8	7	7	0	12	16	8	1	10	7	0	0	0	7	4	6	0	10	8	7	8	7	7
BAGFS	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
BIMAS	11	11	11	11	0	8	8	11	1	8	12	7	0	0	0	5	9	0	8	10	11	10	11	13
BRSAN	5	5	5	5	0	5	5	5	1	5	5	4	0	0	5	0	5	0	5	5	5	5	5	5
BRYAT	11	9	9	12	0	7	7	9	2	8	9	6	0	0	9	5	0	0	8	10	11	10	12	11
BUCIM	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	9	10	8	9	0	11	11	10	1	16	8	10	0	0	8	5	8	0	0	12	10	12	9	9
CEMTS	11	13	10	11	0	9	9	13	1	12	10	8	0	0	10	5	10	0	12	0	12	16	11	11
CIMSA	12	10	11	18	0	8	8	10	1	10	14	7	0	0	11	5	11	0	10	12	0	12	18	13
DOAS	15	17	14	11	0	9	11	17	1	12	11	8	0	0	10	5	10	0	12	16	12	0	11	12
DOHOL	13	10	11	20	0	8	8	10	1	9	14	7	0	0	11	5	12	0	9	11	18	11	0	15
ECILC	14	11	12	15	0	8	8	11	1	9	15	7	0	0	13	5	11	0	9	11	13	12	15	0
ECZYT	12	10	11	18	0	8	8	10	1	10	14	7	0	0	11	5	11	0	10	12	20	12	18	13
EGEEN	13	10	11	19	0	8	8	10	1	9	14	7	0	0	11	5	13	0	9	11	18	11	19	14
EKGYO	14	11	12	15	0	8	8	11	1	9	15	7	0	0	13	5	11	0	9	11	13	12	15	20
ENJSA	13	10	11	19	0	8	8	10	1	9	14	7	0	0	11	5	13	0	9	11	18	11	19	14
ENKAI	8	10	8	8	0	14	12	10	1	14	8	11	0	0	8	5	7	0	14	10	8	10	8	8
EREGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FROTO	12	10	11	14	0	8	8	10	1	11	12	7	0	0	11	5	11	0	12	12	15	12	14	12
GARAN	16	14	16	11	0	9	11	14	1	8	12	8	0	0	11	5	9	0	8	10	11	14	11	12
GLYHO	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1
GSDHO	10	12	10	10	0	9	9	12	1	9	12	8	0	0	11	5	9	0	9	15	10	11	10	10
GUBRF	13	10	11	19	0	8	8	10	1	9	14	7	0	0	11	5	13	0	9	11	18	11	19	14
HALKB	3	3	3	3	0	3	3	3	1	3	3	3	0	0	3	3	3	0	3	3	3	3	3	3
HEKTS	16	14	15	12	0	8	10	14	1	10	12	7	0	0	13	5	10	0	10	12	14	16	12	15
IPEKE	10	12	10	10	0	9	9	12	1	9	11	8	0	0	11	5	9	0	9	13	10	11	10	10
ISCTR	17	14	20	11	0	8	10	14	1	8	12	7	0	0	11	5	9	0	8	10	11	14	11	12
ISDMR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISMEN	13	10	11	19	0	8	8	10	1	9	14	7	0	0	11	5	13	0	9	11	18	11	19	14
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	8	8	7	8	0	10	12	8	1	12	7	12	0	0	7	4	7	0	14	10	10	10	8	8
KCHOL	14	17	14	10	0	9	11	17	1	10	11	8	0	0	10	5	9	0	10	13	10	17	10	11
KONYA	13	10	11	16	0	8	8	10	1	9	13	7	0	0	11	5	16	0	9	11	15	11	16	13
KORDS	9	10	8	9	0	15	11	10	1	17	8	10	0	0	8	5	8	0	15	12	11	12	9	9
KOZAA	9	10	8	9	0	13	11	10	1	20	8	10	0	0	8	5	8	0	16	12	10	12	9	9
KOZAL	13	10	11	17	0	8	8	10	1	9	12	7	0	0	11	5	13	0	9	11	16	11	17	14
KRDMD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAVI	18	14	17	12	0	8	10	14	1	10	12	7	0	0	11	5	10	0	10	12	14	16	12	13
MGROS	14	11	12	15	0	8	8	11	1	9	13	7	0	0	13	5	11	0	9	11	13	12	15	18
ODAS	2	2	2	2	0	2	2	2	0	2	2	2	1	0	2	2	2	0	2	2	2	2	2	2
OTKAR	13	10	11	16	0	8	8	10	1	9	13	7	0	0	11	5	13	0	10	11	15	11	16	13
OYAKC	9	7	7	10	0	7	7	7	1	8	7	6	0	0	7	11	9	0	8	8	8	8	10	10
PETKM	2	4	2	2	0	3	3	4	1	4	2	3	0	0	2	1	1	0	4	4	2	4	2	2
PGSUS	18	14	15	13	0	8	10	14	1	9	12	7	0	0	13	5	11	0	9	11	12	15	13	16

Appendix C 1 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage

	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KONYA	KORDS	KOZAA
AEFES	12	13	14	13	8	0	12	16	1	10	13	3	16	10	17	0	0	13	0	8	14	13	9	9
AGHOL	10	10	11	10	10	0	10	14	1	12	10	3	14	12	14	0	0	10	0	8	17	10	10	10
AKBNK	11	11	12	11	8	0	11	16	1	10	11	3	15	10	20	0	0	11	0	7	14	11	8	8
AKCNS	18	19	15	19	8	0	14	11	1	10	19	3	12	10	11	0	0	19	0	8	10	16	9	9
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	8	8	8	8	14	0	8	9	1	9	8	3	8	9	8	0	0	8	0	10	9	8	15	13
AKSEN	8	8	8	8	12	0	8	11	1	9	8	3	10	9	10	0	0	8	0	12	11	8	11	11
ALARK	10	10	11	10	10	0	10	14	1	12	10	3	14	12	14	0	0	10	0	8	17	10	10	10
ALBRK	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	1	0	1	1	1	1	1
ARCLK	10	9	9	9	14	0	11	8	1	9	9	3	10	9	8	0	0	9	0	12	10	9	17	20
ASELS	14	14	15	14	8	0	12	12	1	12	14	3	12	11	12	0	0	14	0	7	11	13	8	8
ASUZU	7	7	7	7	11	0	7	8	1	8	7	3	7	8	7	0	0	7	0	12	8	7	10	10
BAGFS	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	11	11	13	11	8	0	11	11	1	11	11	3	13	11	11	0	0	11	0	7	10	11	8	8
BRSAN	5	5	5	5	5	0	5	5	1	5	5	3	5	5	5	0	0	5	0	4	5	5	5	5
BRYAT	11	13	11	13	7	0	11	9	1	9	13	3	10	9	9	0	0	13	0	7	9	16	8	8
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	10	9	9	9	14	0	12	8	1	9	9	3	10	9	8	0	0	9	0	14	10	9	15	16
CEMTS	12	11	11	11	10	0	12	10	1	15	11	3	12	13	10	0	0	11	0	10	13	11	12	12
CIMSA	20	18	13	18	8	0	15	11	1	10	18	3	14	10	11	0	0	18	0	10	10	15	11	10
DOAS	12	11	12	11	10	0	12	14	1	11	11	3	16	11	14	0	0	11	0	10	17	11	12	12
DOHOL	18	19	15	19	8	0	14	11	1	10	19	3	12	10	11	0	0	19	0	8	10	16	9	9
ECILC	13	14	20	14	8	0	12	12	1	10	14	3	15	10	12	0	0	14	0	8	11	13	9	9
ECZYT	0	18	13	18	8	0	15	11	1	10	18	3	14	10	11	0	0	18	0	10	10	15	11	10
EGEEN	18	0	14	20	8	0	14	11	1	10	20	3	12	10	11	0	0	20	0	8	10	17	9	9
EKGYO	13	14	0	14	8	0	12	12	1	10	14	3	15	10	12	0	0	14	0	8	11	13	9	9
ENJSA	18	20	14	0	8	0	14	11	1	10	20	3	12	10	11	0	0	20	0	8	10	17	9	9
ENKAI	8	8	8	8	0	0	9	9	1	9	8	3	8	9	8	0	0	8	0	10	10	8	15	14
EREGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0	0
FROTO	15	14	12	14	9	0	0	11	1	10	14	3	13	10	11	0	0	14	0	9	10	14	10	11
GARAN	11	11	12	11	9	0	11	0	1	10	11	3	15	10	16	0	0	11	0	7	14	11	8	8
GLYHO	1	1	1	1	1	0	1	1	0	1	1	3	1	1	1	0	0	1	0	1	1	1	1	1
GSDHO	10	10	10	10	9	0	10	10	1	0	10	3	10	16	10	0	0	10	0	8	11	10	9	9
GUBRF	18	20	14	20	8	0	14	11	1	10	0	3	12	10	11	0	0	20	0	8	10	17	9	9
HALKB	3	3	3	3	3	0	3	3	3	3	3	0	3	3	3	0	0	3	0	3	3	3	3	3
HEKTS	14	12	15	12	8	0	13	15	1	10	12	3	0	10	15	0	0	12	0	10	14	12	11	10
IPEKE	10	10	10	10	9	0	10	10	1	16	10	3	10	0	10	0	0	10	0	8	11	10	9	9
ISCTR	11	11	12	11	8	0	11	16	1	10	11	3	15	10	0	0	0	11	0	7	14	11	8	8
ISDMR	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
ISGYO	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ISMEN	18	20	14	20	8	0	14	11	1	10	20	3	12	10	11	0	0	0	0	8	10	17	9	9
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	10	8	8	8	10	0	9	7	1	8	8	3	10	8	7	0	0	8	0	0	8	8	13	12
KCHOL	10	10	11	10	10	0	10	14	1	11	10	3	14	11	14	0	0	10	0	8	0	10	10	10
KONYA	15	17	13	17	8	0	14	11	1	10	17	3	12	10	11	0	0	17	0	8	10	0	9	9
KORDS	11	9	9	9	15	0	10	8	1	9	9	3	11	9	8	0	0	9	0	13	10	9	0	17
KOZAA	10	9	9	9	14	0	11	8	1	9	9	3	10	9	8	0	0	9	0	12	10	9	17	0
KOZAL	16	18	14	18	8	0	13	11	1	10	18	3	12	10	11	0	0	18	0	8	10	15	9	9
KRDMD	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0	0
MAVI	14	12	13	12	8	0	13	16	1	10	12	3	18	10	17	0	0	12	0	10	14	12	11	10
MGROS	13	14	18	14	8	0	12	12	1	10	14	3	15	10	12	0	0	14	0	8	11	13	9	9
ODAS	2	2	2	2	2	0	2	2	5	2	2	5	2	2	2	0	0	2	0	2	2	2	2	2
OTKAR	15	17	13	17	9	0	17	11	1	10	17	3	12	10	11	0	0	17	0	8	10	17	9	9
OYAKC	8	9	10	9	7	0	8	7	1	7	9	3	8	7	7	0	0	9	0	7	7	9	8	8
PETKM	2	2	2	2	4	9	2	2	0	3	2	1	2	3	2	9	1	2	0	3	4	2	4	4
PGSUS	12	13	16	13	8	0	12	15	1	10	13	3	18	10	15	0	0	13	0	8	14	13	9	9

Appendix C 2 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
AEFES	13	0	18	14	2	13	9	2	18	14	15	12	14	3	0	14	14	13	18	12	10	11	7	11	13	3	14	0	3	0	5	13	2
AGHOL	10	0	14	11	2	10	7	4	14	17	17	12	17	3	0	11	17	10	14	10	7	10	9	10	10	3	17	0	3	0	5	16	2
AKBNK	11	0	17	12	2	11	7	2	15	14	14	12	14	3	0	12	14	11	15	13	7	11	7	14	11	3	14	0	3	0	5	16	2
AKCNS	17	0	12	15	2	16	10	2	13	10	11	12	10	3	0	15	10	17	13	11	9	12	7	12	19	3	10	0	3	0	5	10	2
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	8	0	8	8	2	8	7	3	8	10	9	8	10	3	0	8	10	8	8	8	5	8	5	8	8	3	10	0	3	0	5	9	2
AKSEN	8	0	10	8	2	8	7	3	10	12	11	8	12	3	0	8	12	8	10	8	5	8	5	8	8	3	12	0	3	0	5	11	2
ALARK	10	0	14	11	2	10	7	4	14	17	17	12	17	3	0	11	17	10	14	10	7	10	9	10	10	3	17	0	3	0	5	16	2
ALBRK	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2	1	1	1	1	1	1	0	1	0	2	1	0
ARCLK	9	0	10	9	2	9	8	4	9	10	12	8	10	3	0	9	10	9	9	8	6	8	5	8	9	3	10	0	3	0	5	10	2
ASELS	12	0	12	13	2	13	7	2	12	11	11	16	11	3	0	13	11	14	12	11	7	12	8	12	14	3	11	0	3	0	5	12	2
ASUZU	7	0	7	7	2	7	6	3	7	9	8	7	9	3	0	7	9	7	7	7	5	7	5	7	7	3	9	0	3	0	5	8	2
BAGFS	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	11	0	11	13	2	11	7	2	13	10	10	16	10	3	0	12	10	11	13	11	7	11	8	11	11	3	10	0	3	0	5	10	2
BRSAN	5	0	5	5	2	5	11	1	5	5	5	5	5	3	0	5	5	5	5	5	3	5	3	5	5	3	5	2	3	0	3	5	2
BRYAT	13	0	10	11	2	13	9	1	11	9	10	9	9	3	0	11	9	11	11	9	10	9	7	9	13	3	9	0	3	0	7	9	2
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	9	0	10	9	2	10	8	4	9	10	12	8	10	3	0	9	10	9	9	8	6	9	5	8	9	3	10	0	3	0	5	10	2
CEMTS	11	0	12	11	2	11	8	4	11	13	15	10	13	3	0	11	13	11	11	10	8	10	7	10	11	3	13	0	3	0	5	13	2
CIMSA	16	0	14	13	2	15	8	2	12	10	13	12	10	3	0	13	10	15	12	11	8	12	7	12	18	3	10	0	3	0	5	10	2
DOAS	11	0	16	12	2	11	8	4	15	17	19	11	17	3	0	12	17	11	15	10	8	10	7	10	11	3	17	0	3	0	5	16	2
DOHOL	17	0	12	15	2	16	10	2	13	10	11	12	10	3	0	15	10	17	13	11	9	12	7	12	19	3	10	0	3	0	5	10	2
ECILC	14	0	13	18	2	13	10	2	16	11	12	17	11	3	0	17	11	15	16	11	9	11	7	12	14	3	11	0	3	0	5	12	2
ECZYT	16	0	14	13	2	15	8	2	12	10	13	12	10	3	0	13	10	15	12	11	8	12	7	12	18	3	10	0	3	0	5	10	2
EGEEN	18	0	12	14	2	17	9	2	13	10	11	12	10	3	0	14	10	16	13	11	9	12	7	12	20	3	10	0	3	0	5	10	2
EKGYO	14	0	13	18	2	13	10	2	16	11	12	17	11	3	0	17	11	15	16	11	9	11	7	12	14	3	11	0	3	0	5	12	2
ENJSA	18	0	12	14	2	17	9	2	13	10	11	12	10	3	0	14	10	16	13	11	9	12	7	12	20	3	10	0	3	0	5	10	2
ENKAI	8	0	8	8	2	9	7	4	8	11	10	8	11	3	0	8	11	8	8	8	5	10	5	8	8	3	11	0	3	0	5	10	2
EREGL	0	18	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0
FROTO	13	0	13	12	2	17	8	2	12	10	12	11	10	3	0	12	10	13	12	11	8	14	7	11	14	3	10	0	3	0	5	10	2
GARAN	11	0	16	12	2	11	7	2	15	17	14	12	16	3	0	12	16	11	15	12	7	11	7	11	11	3	16	0	3	0	5	13	2
GLYHO	1	0	1	1	5	1	1	0	1	1	1	1	1	3	0	1	1	1	1	1	1	1	1	1	1	3	1	2	3	0	1	1	5
GSDHO	10	0	10	10	2	10	7	3	10	11	11	11	11	3	0	10	11	10	10	10	7	10	8	10	10	3	11	0	3	0	5	11	2
GUBRF	18	0	12	14	2	17	9	2	13	10	11	12	10	3	0	14	10	16	13	11	9	12	7	12	20	3	10	0	3	0	5	10	2
HALKB	3	0	3	3	5	3	3	1	3	3	3	3	3	7	0	3	3	3	3	3	4	3	5	3	3	9	3	0	17	0	3	3	5
HEKTS	12	0	18	15	2	12	8	2	18	14	17	14	14	3	0	14	14	12	18	11	8	11	7	11	12	3	14	0	3	0	5	13	2
IPEKE	10	0	10	10	2	10	7	3	10	11	11	11	11	3	0	10	11	10	10	10	7	10	8	10	10	3	11	0	3	0	5	11	2
ISCTR	11	0	17	12	2	11	7	2	15	14	14	12	14	3	0	12	14	11	15	13	7	11	7	14	11	3	14	0	3	0	5	16	2
ISDMR	0	19	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0
ISGYO	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ISMEN	18	0	12	14	2	17	9	2	13	10	11	12	10	3	0	14	10	16	13	11	9	12	7	12	20	3	10	0	3	0	5	10	2
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	8	0	10	8	2	8	7	3	8	8	11	7	8	3	0	8	8	8	8	7	6	7	5	7	8	3	8	0	3	0	5	8	2
KCHOL	10	0	14	11	2	10	7	4	14	17	17	11	17	3	0	11	17	10	14	10	7	11	7	10	10	3	17	0	3	0	5	16	2
KONYA	15	0	12	13	2	17	9	2	13	10	11	11	10	3	0	13	10	15	13	11	9	12	7	11	17	3	10	0	3	0	5	10	2
KORDS	9	0	11	9	2	9	8	4	9	10	13	8	10	3	0	9	10	9	9	8	6	8	5	8	9	3	10	0	3	0	5	10	2
KOZAA	9	0	10	9	2	9	8	4	9	10	12	8	10	3	0	9	10	9	9	8	6	8	5	8	9	3	10	0	3	0	5	10	2
KOZAL	0	0	12	14	2	15	9	2	13	10	11	12	10	3	0	14	10	14	13	11	9	11	7	13	18	3	10	0	3	0	5	10	2
KRDMD	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0
MAVI	12	0	0	13	2	12	8	2	16	14	17	12	14	3	0	13	14	12	16	12	8	11	7	11	12	3	14	0	3	0	5	13	2
MGROS	14	0	13	0	2	13	10	2	16	11	12	15	11	3	0	19	11	15	16	11	9	11	7	12	14	3	11	0	3	0	5	11	2
ODAS	2	0	2	2	0	2	2	0	2	2	2	2	2	7	0	2	2	2	2	2	3	2	3	2	2	5	2	0	5	0	2	2	16
OTKAR	15	0	12	13	2	0	9	2	13	10	11	11	10	3	0	13	10	15	13	11	9	13	7	11	17	3	10	0	3	0	5	10	2
OYAKC	9	0	8	10	2	9	0	1	9	7	8	7	7	3	0	10	7	10	9	7	7	7	5	7	9	3	7	0	3	0	5	7	2
PETKM	2	9	2	2	0	2	1	0	2	4	4	2	4	1	0	2	4	2	2	2	1	2	1	2	2	1	4	0	1	11	1	4	0
PGSUS	13	0	16	16	2	13	9	2	0	14	15	14	14	3	0	15	14	13	20	11	10	11	7	11	13	3	14	0	3	0	5	13	2

Appendix C 3 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

Appendix C 4 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKONS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL	ECILC
SAHOL	14	17	14	10	0	10	12	17	1	10	11	9	0	0	10	5	9	0	10	13	10	17	10	11
SASA	15	17	14	11	0	9	11	17	1	12	11	8	0	0	10	5	10	0	12	15	13	19	11	12
SELEC	12	12	12	12	0	8	8	12	1	8	16	7	0	0	16	5	9	0	8	10	12	11	12	17
SISE	14	17	14	10	0	10	12	17	1	10	11	9	0	0	10	5	9	0	10	13	10	17	10	11
SKBNK	3	3	3	3	0	3	3	3	1	3	3	3	0	0	3	3	3	0	3	3	3	3	3	3
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	14	11	12	15	0	8	8	11	1	9	13	7	0	0	12	5	11	0	9	11	13	12	15	17
TAVHL	14	17	14	10	0	10	12	17	1	10	11	9	0	0	10	5	9	0	10	13	10	17	10	11
TCELL	13	10	11	17	0	8	8	10	1	9	14	7	0	0	11	5	11	0	9	11	15	11	17	15
THYAO	18	14	15	13	0	8	10	14	1	9	12	7	0	0	13	5	11	0	9	11	12	15	13	16
TKFEN	12	10	13	11	0	8	8	10	1	8	11	7	0	0	11	5	9	0	8	10	11	10	11	11
TKNSA	10	7	7	9	0	5	5	7	2	6	7	5	0	0	7	3	10	0	6	8	8	8	9	9
TOASO	11	10	11	12	0	8	8	10	1	8	12	7	0	0	11	5	9	0	9	10	12	10	12	11
TSKB	7	9	7	7	0	5	5	9	1	5	8	5	0	0	8	3	7	0	5	7	7	7	7	7
TTKOM	11	10	14	12	0	8	8	10	1	8	12	7	0	0	11	5	9	0	8	10	12	10	12	12
TTRAK	13	10	11	19	0	8	8	10	1	9	14	7	0	0	11	5	13	0	9	11	18	11	19	14
TUKAS	3	3	3	3	0	3	3	3	1	3	3	3	0	0	3	3	3	0	3	3	3	3	3	3
TUPRS	14	17	14	10	0	10	12	17	1	10	11	9	0	0	10	5	9	0	10	13	10	17	10	11
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
VAKBN	3	3	3	3	0	3	3	3	1	3	3	3	0	0	3	3	3	0	3	3	3	3	3	3
VESBE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESTL	5	5	5	5	0	5	5	5	2	5	5	5	0	0	5	3	7	0	5	5	5	5	5	5
YKBNK	13	16	16	10	0	9	11	16	1	10	12	8	0	0	10	5	9	0	10	13	10	16	10	12
ZOREN	2	2	2	2	0	2	2	2	0	2	2	2	1	0	2	2	2	0	2	2	2	2	2	2

Appendix C 5 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

	ECYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KONYA	KORDS	KOZAA
SAHOL	10	10	11	10	11	0	10	17	1	11	10	3	14	11	14	0	0	10	0	8	17	10	10	10
SASA	13	11	12	11	10	0	12	14	1	11	11	3	17	11	14	0	0	11	0	11	17	11	13	12
SELEC	12	12	17	12	8	0	11	12	1	11	12	3	14	11	12	0	0	12	0	7	11	11	8	8
SISE	10	10	11	10	11	0	10	16	1	11	10	3	14	11	14	0	0	10	0	8	17	10	10	10
SKBNK	3	3	3	3	3	0	3	3	3	3	3	7	3	3	3	0	0	3	0	3	3	3	3	3
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	13	14	17	14	8	0	12	12	1	10	14	3	14	10	12	0	0	14	0	8	11	13	9	9
TAVHL	10	10	11	10	11	0	10	16	1	11	10	3	14	11	14	0	0	10	0	8	17	10	10	10
TCELL	15	16	15	16	8	0	13	11	1	10	16	3	12	10	11	0	0	16	0	8	10	15	9	9
THYAO	12	13	16	13	8	0	12	15	1	10	13	3	18	10	15	0	0	13	0	8	14	13	9	9
TKFEN	11	11	11	11	8	0	11	12	1	10	11	3	11	10	13	0	0	11	0	7	10	11	8	8
TKNSA	8	9	9	9	5	0	8	7	1	7	9	4	8	7	7	0	0	9	0	6	7	9	6	6
TOASO	12	12	11	12	10	0	14	11	1	10	12	3	11	10	11	0	0	12	0	7	11	12	8	8
TSKB	7	7	7	7	5	0	7	7	1	8	7	5	7	8	7	0	0	7	0	5	7	7	5	5
TTKOM	12	12	12	12	8	0	11	11	1	10	12	3	11	10	14	0	0	12	0	7	10	11	8	8
TTRAK	18	20	14	20	8	0	14	11	1	10	20	3	12	10	11	0	0	20	0	8	10	17	9	9
TUKAS	3	3	3	3	3	0	3	3	3	3	3	9	3	3	3	0	0	3	0	3	3	3	3	3
TUPRS	10	10	11	10	11	0	10	16	1	11	10	3	14	11	14	0	0	10	0	8	17	10	10	10
ULKER	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	3	3	3	3	3	0	3	3	3	3	3	17	3	3	3	0	0	3	0	3	3	3	3	3
VESBE	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	13	1	0	0	0	0	0	0	0
VESTL	5	5	5	5	5	0	5	5	1	5	5	3	5	5	5	0	0	5	0	5	5	5	5	5
YKBNK	10	10	12	10	10	0	10	13	1	11	10	3	13	11	16	0	0	10	0	8	16	10	10	10
ZOREN	2	2	2	2	2	0	2	2	5	2	2	5	2	2	2	0	0	2	0	2	2	2	2	2

Appendix C 6 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
SAHOL	10	0	14	11	2	10	7	4	14	0	17	11	19	3	0	11	19	10	14	10	7	10	7	10	10	3	19	0	3	0	5	16	2
SASA	11	0	17	12	2	11	8	4	15	17	0	11	17	3	0	12	17	11	15	10	8	10	7	10	11	3	17	0	3	0	5	16	2
SELEC	12	0	12	15	2	11	7	2	14	11	11	0	11	3	0	14	11	12	14	11	7	11	8	12	12	3	11	0	3	0	5	12	2
SISE	10	0	14	11	2	10	7	4	14	19	17	11	0	3	0	11	20	10	14	10	7	10	7	10	10	3	20	0	3	0	5	16	2
SKBNK	3	0	3	3	7	3	3	1	3	3	3	3	3	0	0	3	3	3	3	3	4	3	8	3	3	7	3	0	7	0	3	3	7
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	14	0	13	19	2	13	10	2	15	11	12	14	11	3	0	0	11	15	15	11	9	11	7	12	14	3	11	0	3	0	5	11	2
TAVHL	10	0	14	11	2	10	7	4	14	19	17	11	20	3	0	11	0	10	14	10	7	10	7	10	10	3	20	0	3	0	5	16	2
TCELL	14	0	12	15	2	15	10	2	13	10	11	12	10	3	0	15	10	0	13	11	9	13	7	15	16	3	10	0	3	0	5	10	2
THYAO	13	0	16	16	2	13	9	2	20	14	15	14	14	3	0	15	14	13	0	11	10	11	7	11	13	3	14	0	3	0	5	13	2
TKFEN	11	0	12	11	2	11	7	2	11	10	10	11	10	3	0	11	10	11	11	0	7	11	7	12	11	3	10	0	3	0	5	11	2
TKNSA	9	0	8	9	3	9	7	1	10	7	8	7	7	4	0	9	7	9	10	7	0	7	8	7	9	4	7	0	4	0	6	7	3
TOASO	11	0	11	11	2	13	7	2	11	10	10	11	10	3	0	11	10	13	11	11	7	0	7	12	12	3	10	0	3	0	5	10	2
TSKB	7	0	7	7	3	7	5	1	7	7	7	8	7	8	0	7	7	7	7	7	8	7	0	7	7	5	7	0	5	0	5	7	3
TTKOM	13	0	11	12	2	11	7	2	11	10	10	12	10	3	0	12	10	15	11	12	7	12	7	0	12	3	10	0	3	0	5	13	2
TTRAK	18	0	12	14	2	17	9	2	13	10	11	12	10	3	0	14	10	16	13	11	9	12	7	12	0	3	10	0	3	0	5	10	2
TUKAS	3	0	3	3	5	3	3	1	3	3	3	3	3	7	0	3	3	3	3	3	4	3	5	3	3	0	3	0	10	0	3	3	5
TUPRS	10	0	14	11	2	10	7	4	14	19	17	11	20	3	0	11	20	10	14	10	7	10	7	10	10	3	0	0	3	0	5	16	2
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	3	0	3	3	5	3	3	1	3	3	3	3	3	7	0	3	3	3	3	3	4	3	5	3	3	10	3	0	0	0	3	3	5
VESBE	0	13	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESTL	5	0	5	5	2	5	5	1	5	5	5	5	3	0	0	5	5	5	5	5	6	5	5	5	5	3	5	0	3	0	0	5	2
YKBNK	10	0	13	11	2	10	7	4	13	16	16	12	16	3	0	11	16	10	13	11	7	10	7	13	10	3	16	0	3	0	5	0	2
ZOREN	2	0	2	2	16	2	2	0	2	2	2	2	2	7	0	2	2	2	2	2	3	2	3	2	2	5	2	0	5	0	2	2	0

Appendix C 7 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Average Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL
AEFES	0	22	21	23	1	21	22	22	11	21	22	22	8	5	19	13	20	7	21	22	23	22	23
AGHOL	22	0	22	22	1	21	22	23	11	21	22	22	8	5	19	13	19	7	21	23	22	23	22
AKBNK	21	22	0	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
AKCNS	23	22	21	0	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
AKFGY	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AKSA	21	21	20	21	1	0	21	21	11	20	21	21	8	5	18	13	19	7	20	21	21	21	21
AKSEN	22	22	21	22	1	21	0	22	11	20	22	23	8	5	18	13	19	7	20	22	22	22	22
ALARK	22	23	22	22	1	21	22	0	11	21	22	22	8	5	19	13	19	7	21	23	22	23	22
ALBRK	11	11	11	11	1	11	11	11	0	11	11	11	8	5	11	11	11	7	11	11	11	11	11
ARCLK	21	21	20	21	1	20	20	21	11	0	20	20	8	5	20	13	20	7	22	21	21	21	21
ASELS	22	22	21	23	1	21	22	22	11	20	0	22	8	5	18	13	19	7	20	22	23	22	23
ASUZU	22	22	21	22	1	21	23	22	11	20	22	0	8	5	18	13	19	7	20	22	22	22	22
BAGFS	8	8	8	8	1	8	8	8	8	8	8	8	0	5	8	8	8	7	8	8	8	8	8
BERA	5	5	5	5	1	5	5	5	5	5	5	5	5	0	5	5	5	4	5	5	5	5	5
BIMAS	19	19	18	19	1	18	18	19	11	20	18	18	8	5	0	13	19	7	20	19	19	19	19
BRSAN	13	13	13	13	1	13	13	13	11	13	13	13	8	5	13	0	13	7	13	13	13	13	13
BRYAT	20	19	18	20	1	19	19	19	11	20	19	19	8	5	19	13	0	7	20	19	20	19	20
BUCIM	7	7	7	7	1	7	7	7	7	7	7	7	7	4	7	7	7	0	7	7	7	7	7
CCOLA	21	21	20	21	1	20	20	21	11	22	20	20	8	5	20	13	20	7	0	21	21	21	21
CEMTS	22	23	22	22	1	21	22	23	11	21	22	22	8	5	19	13	19	7	21	0	22	24	22
CIMSA	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	0	22	24
DOAS	22	23	22	22	1	21	22	23	11	21	22	22	8	5	19	13	19	7	21	24	22	0	22
DOHOL	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	0
ECILC	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
ECZYT	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
EGEEN	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
EKGYO	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
ENJSA	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
ENKAI	19	20	20	19	1	19	19	20	11	20	19	19	8	5	18	13	18	7	20	20	19	20	19
EREGL	14	14	13	14	1	14	14	14	11	14	14	14	8	5	14	12	14	7	14	14	14	14	14
FROTO	23	22	21	23	1	21	22	22	11	21	22	22	8	5	19	13	20	7	21	22	23	22	23
GARAN	21	21	21	21	1	20	21	21	11	19	21	21	8	5	17	13	18	7	19	21	21	21	21
GLYHO	9	9	9	9	1	9	9	9	8	9	9	9	8	5	9	9	9	7	9	9	9	9	9
GSDHO	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
GUBRF	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
HALKB	13	13	13	13	1	13	13	13	11	13	13	13	8	5	13	12	13	7	13	13	13	13	13
HEKTS	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
IPEKE	18	18	18	19	1	18	18	18	11	18	19	18	8	5	17	13	17	7	18	18	19	18	19
ISCTR	21	22	24	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
ISDMR	14	14	13	14	1	14	14	14	11	14	14	14	8	5	14	12	14	7	14	14	14	14	14
ISGYO	8	8	8	8	1	8	8	8	8	8	8	8	8	5	8	8	8	7	8	8	8	8	8
ISMEN	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
IZMDC	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
KARSN	17	16	15	17	1	16	16	16	11	17	16	16	8	5	17	12	17	7	17	16	17	16	17
KCHOL	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
KONYA	21	20	19	22	1	20	20	20	11	21	21	20	8	5	19	13	20	7	21	20	22	20	22
KORDS	20	20	19	20	1	19	19	20	11	21	19	19	8	5	20	13	20	7	21	20	20	20	20
KOZAA	21	21	20	21	1	20	20	21	11	22	20	20	8	5	20	13	20	7	22	21	21	21	21
KOZAL	22	21	20	22	1	20	21	21	11	20	21	21	8	5	18	13	19	7	20	21	22	21	22
KRDMD	14	14	13	14	1	14	14	14	11	14	14	14	8	5	14	12	14	7	14	14	14	14	14
MAVI	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
MGROS	22	22	21	22	1	20	21	22	11	22	21	21	8	5	20	13	20	7	22	22	22	22	22
ODAS	12	12	12	12	1	12	12	12	11	12	12	12	8	5	12	12	12	7	12	12	12	12	12
OTKAR	23	22	21	24	1	21	22	22	11	21	23	22	8	5	19	13	20	7	21	22	24	22	24
OYAKC	18	17	16	18	1	17	17	17	11	18	17	17	8	5	18	13	18	7	18	17	18	17	18
PETKM	18	18	17	18	1	18	18	18	11	18	18	18	8	5	18	13	18	7	18	18	18	18	18
PGSUS	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23

Appendix D 1 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage

	ECILC	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KONYA
AEFES	23	23	23	23	23	19	14	23	21	9	21	23	13	23	18	21	14	8	23	2	17	21	21
AGHOL	23	22	22	23	22	20	14	22	21	9	22	22	13	23	18	22	14	8	22	2	16	22	20
AKBNK	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	24	13	8	21	2	15	22	19
AKCNS	23	24	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
AKFGY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
AKSA	21	21	21	21	21	19	14	21	20	9	20	21	13	21	18	20	14	8	21	2	16	20	20
AKSEN	22	22	22	22	22	19	14	22	21	9	21	22	13	22	18	21	14	8	22	2	16	21	20
ALARK	23	22	22	23	22	20	14	22	21	9	22	22	13	23	18	22	14	8	22	2	16	22	20
ALBRK	11	11	11	11	11	11	11	11	11	8	11	11	11	11	11	11	11	8	11	2	11	11	11
ARCLK	22	21	21	22	21	20	14	21	19	9	20	21	13	22	18	20	14	8	21	2	17	20	21
ASELS	22	23	23	22	23	19	14	22	21	9	21	23	13	22	19	21	14	8	23	2	16	21	21
ASUZU	22	22	22	22	22	19	14	22	21	9	21	22	13	22	18	21	14	8	22	2	16	21	20
BAGFS	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	2	8	8	8
BERA	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	5	5	5
BIMAS	20	19	19	20	19	18	14	19	17	9	18	19	13	20	17	18	14	8	19	2	17	18	19
BRSAN	13	13	13	13	13	13	12	13	13	9	13	13	12	13	13	13	12	8	13	2	12	13	13
BRYAT	20	20	20	20	20	18	14	20	18	9	18	20	13	20	17	18	14	8	20	2	17	18	20
BUCIM	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	2	7	7	7
CCOLA	22	21	21	22	21	20	14	21	19	9	20	21	13	22	18	20	14	8	21	2	17	20	21
CEMTS	23	22	22	23	22	20	14	22	21	9	22	22	13	23	18	22	14	8	22	2	16	22	20
CIMSA	23	24	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
DOAS	23	22	22	23	22	20	14	22	21	9	22	22	13	23	18	22	14	8	22	2	16	22	20
DOHOL	23	24	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
ECILC	0	23	23	24	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21
ECZYT	23	0	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
EGEEN	23	24	0	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
EKGYO	24	23	23	0	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21
ENJSA	23	24	24	23	0	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
ENKAI	20	19	19	20	19	0	13	19	19	9	20	19	13	20	18	20	13	8	19	2	15	20	19
EREGL	14	14	14	14	14	13	0	14	13	9	13	14	12	14	13	13	20	8	14	2	14	13	14
FROTO	23	23	23	23	23	19	14	0	21	9	21	23	13	23	18	21	14	8	23	2	17	21	21
GARAN	21	21	21	21	21	19	13	21	0	9	21	21	13	21	18	21	13	8	21	2	15	21	19
GLYHO	9	9	9	9	9	9	9	9	9	0	9	9	9	9	9	9	9	8	9	2	9	9	9
GSDHO	22	21	21	22	21	20	13	21	21	9	0	21	13	22	18	22	13	8	21	2	15	22	19
GUBRF	23	24	24	23	24	19	14	23	21	9	21	0	13	23	19	21	14	8	24	2	17	21	22
HALKB	13	13	13	13	13	13	12	13	13	9	13	13	0	13	13	13	12	8	13	2	13	13	13
HEKTS	24	23	23	24	23	20	14	23	21	9	22	23	13	0	18	22	14	8	23	2	17	22	21
IPEKE	18	19	19	18	19	18	13	18	18	9	18	19	13	18	0	18	13	8	19	2	15	18	19
ISCTR	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	0	13	8	21	2	15	22	19
ISDMR	14	14	14	14	14	13	20	14	13	9	13	14	12	14	13	13	0	8	14	2	14	13	14
ISGYO	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	0	8	2	8	8	8
ISMEN	23	24	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	0	2	17	21	22
IZMDC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2
KARSN	17	17	17	17	17	15	14	17	15	9	15	17	13	17	15	15	14	8	17	2	0	15	17
KCHOL	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	0	19
KONYA	21	22	22	21	22	19	14	21	19	9	19	22	13	21	19	19	14	8	22	2	17	19	0
KORDS	21	20	20	21	20	19	14	20	18	9	19	20	13	21	17	19	14	8	20	2	17	19	20
KOZAA	22	21	21	22	21	20	14	21	19	9	20	21	13	22	18	20	14	8	21	2	17	20	21
KOZAL	22	22	22	22	22	18	14	22	20	9	20	22	13	22	17	20	14	8	22	2	17	20	20
KRDMD	14	14	14	14	14	13	17	14	13	9	13	14	12	14	13	13	17	8	14	2	14	13	14
MAVI	24	23	23	24	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21
MGROS	23	22	22	23	22	20	14	22	20	9	21	22	13	23	18	21	14	8	22	2	17	21	21
ODAS	12	12	12	12	12	12	12	12	12	9	12	12	12	12	12	12	12	8	12	2	12	12	12
OTKAR	23	24	24	23	24	19	14	23	21	9	21	24	13	23	19	21	14	8	24	2	17	21	22
OYAKC	18	18	18	18	18	16	14	18	16	9	16	18	13	18	16	16	14	8	18	2	17	16	18
PETKM	18	18	18	18	18	17	14	18	17	9	17	18	13	18	17	17	14	8	18	2	16	17	18
PGSUS	24	23	23	24	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21

Appendix D 2 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage (Continued)

	KORDS	KOZAA	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
AEFES	20	21	22	14	23	22	12	23	18	18	23	21	23	21	21	12	1	22	21	20	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
AGHOL	20	21	21	14	23	22	12	22	17	18	23	22	23	22	22	12	1	21	22	19	23	16	15	18	14	17	22	12	22	8	13	14	12	22	12
AKBNK	19	20	20	13	22	21	12	21	16	17	22	22	22	22	22	12	1	20	22	18	22	16	14	18	14	19	21	12	22	8	13	13	12	24	12
AKCNS	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
AKFGY	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AKSA	19	20	20	14	21	20	12	21	17	18	21	20	21	20	20	12	1	20	20	19	21	16	15	18	14	17	21	12	20	8	13	14	12	20	12
AKSEN	19	20	21	14	22	21	12	22	17	18	22	21	22	21	21	12	1	21	21	19	22	16	15	18	14	17	22	12	21	8	13	14	12	21	12
ALARK	20	21	21	14	23	22	12	22	17	18	23	22	23	22	22	12	1	21	22	19	23	16	15	18	14	17	22	12	22	8	13	14	12	22	12
ALBRK	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	1	11	11	11	11	11	11	11	11	11	11	11	11	8	11	11	11	11	11
ARCLK	21	22	20	14	22	22	12	21	18	18	22	20	22	20	20	12	1	21	20	20	22	16	16	18	14	17	21	12	20	8	13	14	12	20	12
ASELS	19	20	21	14	22	21	12	23	17	18	22	21	22	21	21	12	1	21	21	20	22	16	15	18	14	17	22	12	21	8	13	14	12	21	12
ASUZU	19	20	21	14	22	21	12	22	17	18	22	21	22	21	21	12	1	21	21	19	22	16	15	18	14	17	22	12	21	8	13	14	12	21	12
BAGFS	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
BERA	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
BIMAS	20	20	18	14	20	20	12	19	18	18	20	18	20	18	18	12	1	19	18	19	20	16	16	16	14	16	19	12	18	8	13	14	12	18	12
BRSAN	13	13	13	12	13	13	12	13	13	13	13	13	13	13	13	12	1	13	13	13	13	13	12	13	13	13	13	12	13	8	12	12	12	13	12
BRYAT	20	20	19	14	20	20	12	20	18	18	20	18	20	18	18	12	1	20	18	19	20	16	16	17	14	16	20	12	18	8	13	14	12	18	12
BUCIM	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	1	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
CCOLA	21	22	20	14	22	22	12	21	18	18	22	20	22	20	20	12	1	21	20	20	22	16	16	18	14	17	21	12	20	8	13	14	12	20	12
CEMTS	20	21	21	14	23	22	12	22	17	18	23	22	23	22	22	12	1	21	22	19	23	16	15	18	14	17	22	12	22	8	13	14	12	22	12
CIMSA	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
DOAS	20	21	21	14	23	22	12	22	17	18	23	22	23	22	22	12	1	21	22	19	23	16	15	18	14	17	22	12	22	8	13	14	12	22	12
DOHOL	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
ECILC	21	22	22	14	24	23	12	23	18	18	24	22	24	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12
ECZYT	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
EGEEN	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
EKGYO	21	22	22	14	24	23	12	23	18	18	24	22	24	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12
ENJSA	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
ENKAI	19	20	18	13	20	20	12	19	16	17	20	20	20	20	20	12	1	19	20	18	20	16	14	18	14	17	19	12	20	8	13	13	12	20	12
EREGL	14	14	14	17	14	14	12	14	14	14	14	13	14	13	13	12	1	14	13	14	14	13	14	12	12	13	14	12	13	8	12	16	12	13	12
FROTO	20	21	22	14	23	22	12	23	18	18	23	21	23	21	21	12	1	22	21	20	23	16	16	19	14	17	23	12	21	8	13	14	12	21	12
GARAN	18	19	20	13	21	20	12	21	16	17	21	21	21	21	21	12	1	20	21	18	21	16	14	18	14	17	21	12	21	8	13	13	12	21	12
GLYHO	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	9	9	9	9	9	9	9	9	9	9	9	9	8	9	9	9	9	9
GSDHO	19	20	20	13	22	21	12	21	16	17	22	22	22	22	22	12	1	20	22	18	22	16	14	18	14	17	21	12	22	8	13	13	12	22	12
GUBRF	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
HALKB	13	13	13	12	13	13	12	13	13	13	13	13	13	13	13	12	1	13	13	13	13	13	12	13	12	13	13	12	13	8	19	12	12	13	12
HEKTS	21	22	22	14	24	23	12	23	18	18	24	22	24	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12
IPEKE	17	18	17	13	18	18	12	19	16	17	18	18	18	18	18	12	1	18	18	19	18	16	14	17	14	17	18	12	18	8	13	13	12	18	12
ISCTR	19	20	20	13	22	21	12	21	16	17	22	22	22	22	22	12	1	20	22	18	22	16	14	18	14	19	21	12	22	8	13	13	12	24	12
ISDMR	14	14	14	17	14	14	12	14	14	14	14	13	14	13	13	12	1	14	13	14	14	13	14	12	12	13	14	12	13	8	12	16	12	13	12
ISGYO	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	1	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
ISMEN	20	21	22	14	23	22	12	24	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
IZMDC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
KARSN	17	17	17	14	17	17	12	17	17	16	17	15	17	15	15	12	1	17	15	17	17	15	16	14	13	15	17	12	15	8	13	14	12	15	12
KCHOL	19	20	20	13	22	21	12	21	16	17	22	22	22	22	22	12	1	20	22	18	22	16	14	18	14	17	21	12	22	8	13	13	12	22	12
KONYA	20	21	20	14	21	21	12	22	18	18	21	19	21	19	19	12	1	21	19	21	21	16	16	18	14	17	21	12	19	8	13	14	12	19	12
KORDS	0	21	19	14	21	21	12	20	18	18	21	19	21	19	19	12	1	20	19	19	21	16	16	17	14	16	20	12	19	8	13	14	12	19	12
KOZAA	21	0	20	14	22	22	12	21	18	18	22	20	22	20	20	12	1	21	20	20	22	16	16	18	14	17	21	12	20	8	13	14	12	20	12
KOZAL	19	20	0	14	22	21	12	22	18	17	22	20	22	20	20	12	1	21	20	19	22	16	16	17	14	17	22	12	20	8	13	14	12	20	12
KRDMD	14	14	14	0	14	14	12	14	14	14	14	13	14	13	13	12	1	14	13	14	14	13	14	12	12	13	14	12	13	8	12	15	12	13	12
MAVI	21	22	22	14	0	23	12	23	18	18	24	22	24	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12
MGROS	21	22	21	14	23	0	12	22	18	18	23	21	23	21	21	12	1	23	21	20	23	16	16	18	14	17	22	12	21	8	13	14	12	21	12
ODAS	12	12	12	12	12	12	0	12	12	12	12	12	12	12	12	12	1	12	12	12	12	12	12	12	12	12	12	12	12	8	12	12	12	12	15
OTKAR	20	21	22	14	23	22	12	0	18	18	23	21	23	21	21	12	1	22	21	21	23	16	16	18	14	17	23	12	21	8	13	14	12	21	12
OYAKC	18	18	18	14	18	18	12	18	0	17	18	16	18	16	16	12	1	18	16	18	18	16	16	15	14	16	18	12	16	8	13	14	12	16	12
PETKM	18	18	17	14	18	18	12	18	17	0	18	17	18	17	17	12	1	18	17	18	18	16	15	16	14	16	18	12	17	8	13	14	12	17	12
PGSUS	21	22	22	14	24	23	12	23	18	18	0	22	24	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12

Appendix D 3 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL
SAHOL	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
SASA	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
SELEC	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
SISE	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
SKBNK	12	12	12	12	1	12	12	12	11	12	12	12	8	5	12	12	12	7	12	12	12	12	12
SNGYO	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SOKM	22	21	20	22	1	20	21	21	11	21	21	21	8	5	19	13	20	7	21	21	22	21	22
TAVHL	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
TCELL	20	19	18	21	1	19	19	19	11	20	20	19	8	5	19	13	19	7	20	19	21	19	21
THYAO	23	23	22	23	1	21	22	23	11	22	22	22	8	5	20	13	20	7	22	23	23	23	23
TKFEN	16	16	16	16	1	16	16	16	11	16	16	16	8	5	16	13	16	7	16	16	16	16	16
TKNSA	16	15	14	16	1	15	15	15	11	16	15	15	8	5	16	12	16	7	16	15	16	15	16
TOASO	18	18	18	18	1	18	18	18	11	18	18	18	8	5	16	13	17	7	18	18	18	18	18
TSKB	14	14	14	14	1	14	14	14	11	14	14	14	8	5	14	13	14	7	14	14	14	14	14
TTKOM	17	17	19	17	1	17	17	17	11	17	17	17	8	5	16	13	16	7	17	17	17	17	17
TTRAK	23	22	21	23	1	21	22	22	11	21	22	22	8	5	19	13	20	7	21	22	23	22	23
TUKAS	12	12	12	12	1	12	12	12	11	12	12	12	8	5	12	12	12	7	12	12	12	12	12
TUPRS	21	22	22	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
ULKER	8	8	8	8	1	8	8	8	8	8	8	8	8	5	8	8	8	7	8	8	8	8	8
VAKBN	13	13	13	13	1	13	13	13	11	13	13	13	8	5	13	12	13	7	13	13	13	13	13
VESBE	14	14	13	14	1	14	14	14	11	14	14	14	8	5	14	12	14	7	14	14	14	14	14
VESTL	12	12	12	12	1	12	12	12	11	12	12	12	8	5	12	12	12	7	12	12	12	12	12
YKBNK	21	22	24	21	1	20	21	22	11	20	21	21	8	5	18	13	18	7	20	22	21	22	21
ZOREN	12	12	12	12	1	12	12	12	11	12	12	12	8	5	12	12	12	7	12	12	12	12	12

Appendix D 4 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage (Continued)

	ECILC	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KOMFA
SAHOL	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	22	19
SASA	24	23	23	24	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21
SELEC	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	22	19
SISE	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	22	19
SKBNK	12	12	12	12	12	12	12	12	12	9	12	12	12	12	12	12	12	8	12	2	12	12	12
SNGYO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
SOKM	22	22	22	22	22	19	14	22	20	9	20	22	13	22	18	20	14	8	22	2	17	20	21
TAVHL	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	22	19
TCELL	20	21	21	20	21	18	14	20	18	9	18	21	13	20	19	18	14	8	21	2	17	18	21
THYAO	24	23	23	24	23	20	14	23	21	9	22	23	13	24	18	22	14	8	23	2	17	22	21
TKFEN	16	16	16	16	16	16	13	16	16	9	16	16	13	16	16	16	13	8	16	2	15	16	16
TKNSA	16	16	16	16	16	14	14	16	14	9	14	16	12	16	14	14	14	8	16	2	16	14	16
TOASO	18	18	18	18	18	18	12	19	18	9	18	18	13	18	17	18	12	8	18	2	14	18	18
TSKB	14	14	14	14	14	14	12	14	14	9	14	14	12	14	14	14	12	8	14	2	13	14	14
TTKOM	17	17	17	17	17	17	13	17	17	9	17	17	13	17	17	19	13	8	17	2	15	17	17
TTRAK	23	23	23	23	23	19	14	23	21	9	21	23	13	23	18	21	14	8	23	2	17	21	21
TUKAS	12	12	12	12	12	12	12	12	12	9	12	12	12	12	12	12	12	8	12	2	12	12	12
TUPRS	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	22	13	8	21	2	15	22	19
ULKER	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	2	8	8	8
VAKBN	13	13	13	13	13	13	12	13	13	9	13	13	19	13	13	13	12	8	13	2	13	13	13
VESBE	14	14	14	14	14	13	16	14	13	9	13	14	12	14	13	13	16	8	14	2	14	13	14
VESTL	12	12	12	12	12	12	12	12	12	9	12	12	12	12	12	12	12	8	12	2	12	12	12
YKBNK	22	21	21	22	21	20	13	21	21	9	22	21	13	22	18	24	13	8	21	2	15	22	19
ZOREN	12	12	12	12	12	12	12	12	12	9	12	12	12	12	12	12	12	8	12	2	12	12	12

Appendix D 5 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage (Continued)

	KORDS	KOZAA	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
SAHOL	19	20	20	13	22	21	12	21	16	17	22	0	22	22	22	12	1	20	22	18	22	16	14	18	14	17	21	12	22	8	13	13	12	22	12
SASA	21	22	22	14	24	23	12	23	18	18	24	22	0	22	22	12	1	22	22	20	24	16	16	18	14	17	23	12	22	8	13	14	12	22	12
SELEC	19	20	20	13	22	21	12	21	16	17	22	22	22	0	22	12	1	20	22	18	22	16	14	18	14	17	21	12	22	8	13	13	12	22	12
SISE	19	20	20	13	22	21	12	21	16	17	22	22	22	22	0	12	1	20	24	18	22	16	14	18	14	17	21	12	24	8	13	13	12	22	12
SKBNK	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	0	1	12	12	12	12	12	12	12	13	12	12	12	12	8	12	12	12	12	12
SNGYO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SOKM	20	21	21	14	22	23	12	22	18	18	22	20	22	20	20	12	1	0	20	20	22	16	16	18	14	17	22	12	20	8	13	14	12	20	12
TAVHL	19	20	20	13	22	21	12	21	16	17	22	22	22	22	24	12	1	20	0	18	22	16	14	18	14	17	21	12	24	8	13	13	12	22	12
TCELL	19	20	19	14	20	20	12	21	18	18	20	18	20	18	18	12	1	20	18	0	20	16	16	17	14	20	20	12	18	8	13	14	12	18	12
THYAO	21	22	22	14	24	23	12	23	18	18	24	22	24	22	22	12	1	22	22	20	0	16	16	18	14	17	23	12	22	8	13	14	12	22	12
TKFEN	16	16	16	13	16	16	12	16	16	16	16	16	16	16	16	12	1	16	16	16	16	0	14	15	14	16	16	12	16	8	13	13	12	16	12
TKNSA	16	16	16	14	16	16	12	16	16	15	16	14	16	14	14	12	1	16	14	16	16	14	0	13	13	14	16	12	14	8	12	14	12	14	12
TOASO	17	18	17	12	18	18	12	18	15	16	18	18	18	18	18	12	1	18	18	17	18	15	13	0	14	16	18	12	18	8	13	12	12	18	12
TSKB	14	14	14	12	14	14	12	14	14	14	14	14	14	14	14	13	1	14	14	14	14	14	13	14	0	14	14	12	14	8	12	12	12	14	12
TTKOM	16	17	17	13	17	17	12	17	16	16	17	17	17	17	17	12	1	17	17	20	17	16	14	16	14	0	17	12	17	8	13	13	12	19	12
TTRAK	20	21	22	14	23	22	12	23	18	18	23	21	23	21	21	12	1	22	21	20	23	16	16	18	14	17	0	12	21	8	13	14	12	21	12
TUKAS	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	1	12	12	12	12	12	12	12	12	12	12	0	12	8	12	12	12	12	12
TUPRS	19	20	20	13	22	21	12	21	16	17	22	22	22	22	24	12	1	20	24	18	22	16	14	18	14	17	21	12	0	8	13	13	12	22	12
ULKER	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	1	8	8	8	8	8	8	8	8	8	8	8	8	0	8	8	8	8	8
VAKBN	13	13	13	12	13	13	12	13	13	13	13	13	13	13	13	12	1	13	13	13	13	13	12	13	12	13	13	12	13	8	0	12	12	13	12
VESBE	14	14	14	15	14	14	12	14	14	14	14	13	14	13	13	12	1	14	13	14	14	13	14	12	12	13	14	12	13	8	12	0	12	13	12
VESTL	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	1	12	12	12	12	12	12	12	12	12	12	12	12	8	12	12	0	12	12
YKBNK	19	20	20	13	22	21	12	21	16	17	22	22	22	22	22	12	1	20	22	18	22	16	14	18	14	19	21	12	22	8	13	13	12	0	12
ZOREN	12	12	12	12	12	12	15	12	12	12	12	12	12	12	12	12	1	12	12	12	12	12	12	12	12	12	12	12	12	8	12	12	12	12	0

Appendix D 6 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Single Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKONS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN
AEFES	0	3	5	0	0	0	0	3	0	0	0	0	0	0	0	0
AGHOL	3	0	3	0	0	0	0	5	0	0	0	0	0	0	0	0
AKBNK	5	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0
AKONS	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	1
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	0	0	0	0	0	4	0	0	2	0	4	0	0	0	0
AKSEN	0	0	0	0	0	4	0	0	0	2	0	5	0	0	0	0
ALARK	3	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0
ALBRK	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
ARCLK	0	0	0	0	0	2	2	0	0	0	0	2	0	0	0	0
ASELS	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1
ASUZU	0	0	0	0	0	4	5	0	0	2	0	0	0	0	0	0
BAGFS	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	2
BRSAN	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2	0
BRYAT	0	0	0	3	0	0	0	0	0	0	2	0	0	0	1	1
BUCIM	0	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0
CCOLA	0	0	0	0	0	2	2	0	0	5	0	2	0	0	0	0
CEMTS	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
CIMSA	0	0	0	4	0	0	0	0	0	0	2	0	0	0	1	1
DOAS	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
DOHOL	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
ECILC	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	2
ECZYT	0	0	0	4	0	0	0	0	0	0	2	0	0	0	1	1
EGEEN	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
EKGYO	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	2
ENJSA	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
ENKAI	0	0	0	0	0	2	2	0	0	4	0	2	0	0	0	0
EREGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FROTO	0	0	0	2	0	0	0	0	0	0	3	0	0	0	1	1
GARAN	5	3	5	0	0	0	0	3	0	0	0	0	0	0	0	0
GLYHO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GSDHO	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
GUBRF	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
HALKB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HEKTS	0	0	0	1	0	0	0	0	0	0	1	0	0	0	4	2
IPEKE	0	0	0	2	0	0	0	0	0	0	5	0	0	0	1	1
ISCTR	5	3	5	0	0	0	0	3	0	0	0	0	0	0	0	0
ISDMR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISMEN	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	0	0	0	0	0	4	5	0	0	2	0	5	0	0	0	0
KCHOL	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
KONYA	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
KORDS	0	0	0	0	0	2	2	0	0	4	0	2	0	0	0	0
KOZAA	0	0	0	0	0	2	2	0	0	5	0	2	0	0	0	0
KOZAL	0	0	0	2	0	0	0	0	0	0	3	0	0	0	1	1
KRDMD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAVI	5	3	5	0	0	0	0	3	0	0	0	0	0	0	0	0
MGROS	0	0	0	1	0	0	0	0	0	0	1	0	0	0	3	2
ODAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OTKAR	0	0	0	4	0	0	0	0	0	0	2	0	0	0	1	1
OYAKC	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2	4
PETKM	0	0	0	0	0	2	2	0	0	4	0	2	0	0	0	0
PGSUS	0	0	0	1	0	0	0	0	0	0	1	0	0	0	4	2

Appendix E 1 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage

	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL	ECILC	ECZYT	EGEEN	EKGYO	ENISA	ENKAI	EREGL	FROTO	GARAN
AEFES	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
AGHOL	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
AKBNK	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
AKCNS	3	0	0	0	4	0	3	1	4	3	1	3	0	0	2	0
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0
AKSEN	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0
ALARK	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
ALBRK	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARCLK	0	0	3	0	0	0	0	0	0	0	0	0	4	0	0	0
ASELS	2	0	0	0	2	0	2	1	2	2	1	2	0	0	3	0
ASUZU	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0
BAGFS	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	1	0	0	0	1	0	1	3	1	1	3	1	0	0	1	0
BRSAN	1	0	0	0	1	0	1	2	1	1	2	1	0	0	1	0
BRYAT	0	0	0	0	3	0	3	1	3	3	1	3	0	0	2	0
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
CEMTS	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	3
CIMSA	3	0	0	0	0	0	4	1	3	4	1	4	0	0	2	0
DOAS	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	3
DOHOL	3	0	0	0	4	0	0	1	4	3	1	3	0	0	2	0
ECILC	1	0	0	0	1	0	1	0	1	1	3	1	0	0	1	0
ECZYT	3	0	0	0	3	0	4	1	0	4	1	4	0	0	2	0
EGEEN	3	0	0	0	4	0	3	1	4	0	1	3	0	0	2	0
EKGYO	1	0	0	0	1	0	1	3	1	1	0	1	0	0	1	0
ENISA	3	0	0	0	4	0	3	1	4	3	1	0	0	0	2	0
ENKAI	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
EREGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FROTO	2	0	0	0	2	0	2	1	2	2	1	2	0	0	0	0
GARAN	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	0
GLYHO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GSDHO	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
GUBRF	3	0	0	0	4	0	3	1	4	3	1	3	0	0	2	0
HALKB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HEKTS	1	0	0	0	1	0	1	4	1	1	4	1	0	0	1	0
IPEKE	2	0	0	0	2	0	2	1	2	2	1	2	0	0	3	0
ISCTR	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
ISDMR	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
ISGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
ISMEN	3	0	0	0	4	0	3	1	4	3	1	3	0	0	2	0
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0
KCHOL	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	3
KONYA	3	0	0	0	4	0	3	1	4	3	1	3	0	0	2	0
KORDS	0	0	4	0	0	0	0	0	0	0	0	0	3	0	0	0
KOZAA	0	0	3	0	0	0	0	0	0	0	0	0	4	0	0	0
KOZAL	2	0	0	0	2	0	2	1	2	2	1	2	0	0	3	0
KRDMD	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
MAVI	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
MGROS	1	0	0	0	1	0	1	3	1	1	3	1	0	0	1	0
ODAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OTKAR	3	0	0	0	3	0	4	1	3	4	1	4	0	0	2	0
OYAKC	1	0	0	0	1	0	1	2	1	1	2	1	0	0	1	0
PETKM	0	0	4	0	0	0	0	0	0	0	0	0	4	0	0	0
PGSUS	1	0	0	0	1	0	1	4	1	1	4	1	0	0	1	0

Appendix E 2 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KONYA	KORDS	KOZAA
AEFES	0	3	0	0	0	0	5	0	0	0	0	0	3	0	0	0
AGHOL	0	3	0	0	0	0	3	0	0	0	0	0	3	0	0	0
AKBNK	0	3	0	0	0	0	5	0	0	0	0	0	3	0	0	0
AKCNS	0	0	5	0	1	2	0	0	0	5	0	0	0	5	0	0
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	0	0	0	0	0	0	0	0	0	0	4	0	0	2	2
AKSEN	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	2
ALARK	0	3	0	0	0	0	3	0	0	0	0	0	3	0	0	0
ALBRK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARCLK	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	5
ASELS	0	0	2	0	1	5	0	0	0	2	0	0	0	2	0	0
ASUZU	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	2
BAGFS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	0	0	1	0	4	1	0	0	0	1	0	0	0	1	0	0
BRSAN	0	0	1	0	2	1	0	0	0	1	0	0	0	1	0	0
BRYAT	0	0	3	0	1	2	0	0	0	3	0	0	0	3	0	0
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	5
CEMTS	0	5	0	0	0	0	3	0	0	0	0	0	4	0	0	0
CIMSA	0	0	4	0	1	2	0	0	0	4	0	0	0	4	0	0
DOAS	0	5	0	0	0	0	3	0	0	0	0	0	4	0	0	0
DOHOL	0	0	5	0	1	2	0	0	0	5	0	0	0	5	0	0
ECILC	0	0	1	0	4	1	0	0	0	1	0	0	0	1	0	0
ECZYT	0	0	4	0	1	2	0	0	0	4	0	0	0	4	0	0
EGEEN	0	0	5	0	1	2	0	0	0	5	0	0	0	5	0	0
EKGYO	0	0	1	0	4	1	0	0	0	1	0	0	0	1	0	0
ENJSA	0	0	5	0	1	2	0	0	0	5	0	0	0	5	0	0
ENKAI	0	0	0	0	0	0	0	0	0	0	0	2	0	0	5	4
EREGL	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	0
FROTO	0	0	2	0	1	3	0	0	0	2	0	0	0	2	0	0
GARAN	0	3	0	0	0	0	5	0	0	0	0	0	3	0	0	0
GLYHO	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
GSDHO	0	0	0	0	0	0	3	0	0	0	0	0	4	0	0	0
GUBRF	0	0	0	0	1	2	0	0	0	5	0	0	0	5	0	0
HALKB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HEKTS	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0
IPEKE	0	0	2	0	1	0	0	0	0	2	0	0	0	2	0	0
ISCTR	0	3	0	0	0	0	0	0	0	0	0	0	3	0	0	0
ISDMR	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
ISGYO	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
ISMEN	0	0	5	0	1	2	0	0	0	0	0	0	0	5	0	0
IZMDC	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
KCHOL	0	4	0	0	0	0	3	0	0	0	0	0	0	0	0	0
KONYA	0	0	5	0	1	2	0	0	0	5	0	0	0	0	0	0
KORDS	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	4
KOZAA	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	0
KOZAL	0	0	2	0	1	3	0	0	0	2	0	0	0	2	0	0
KRDMD	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	0
MAVI	0	3	0	0	0	0	5	0	0	0	0	0	3	0	0	0
MGROS	0	0	1	0	3	1	0	0	0	1	0	0	0	1	0	0
ODAS	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
OTKAR	0	0	4	0	1	2	0	0	0	4	0	0	0	4	0	0
OYAKC	0	0	1	0	2	1	0	0	0	1	0	0	0	1	0	0
PETKM	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4	4
PGSUS	0	0	1	0	5	1	0	0	0	1	0	0	0	1	0	0

Appendix E 3 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SMGYO	SODAS
AEFES	0	0	5	0	0	0	0	0	0	3	3	0	3	0	0	0
AGHOL	0	0	3	0	0	0	0	0	0	3	3	0	3	0	0	0
AKBNK	0	0	5	0	0	0	0	0	0	3	3	0	3	0	0	0
AKCNS	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
AKSEN	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
ALARK	0	0	3	0	0	0	0	0	0	3	3	0	3	0	0	0
ALBRK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ARCLK	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
ASELS	3	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
ASUZU	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
BAGFS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	1	0	0	3	0	1	2	0	4	0	0	5	0	0	0	3
BRSAN	1	0	0	2	0	1	4	0	2	0	0	2	0	0	0	2
BRYAT	2	0	0	1	0	3	1	0	1	0	0	1	0	0	0	1
BUCIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
CEMTS	0	0	3	0	0	0	0	0	0	4	5	0	5	0	0	0
CIMSA	2	0	0	1	0	5	1	0	1	0	0	1	0	0	0	1
DOAS	0	0	3	0	0	0	0	0	0	4	5	0	5	0	0	0
DOHOL	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
ECILC	1	0	0	3	0	1	2	0	4	0	0	5	0	0	0	3
ECZYT	2	0	0	1	0	5	1	0	1	0	0	1	0	0	0	1
EGEEN	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
EKGYO	1	0	0	3	0	1	2	0	4	0	0	5	0	0	0	3
ENJSA	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
ENKAI	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
EREGL	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FROTO	3	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
GARAN	0	0	5	0	0	0	0	0	0	3	3	0	3	0	0	0
GLYHO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GSDHO	0	0	3	0	0	0	0	0	0	4	5	0	5	0	0	0
GUBRF	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
HALKB	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0
HEKTS	1	0	0	3	0	1	2	0	5	0	0	4	0	0	0	3
IPEKE	3	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
ISCTR	0	0	5	0	0	0	0	0	0	3	3	0	3	0	0	0
ISDMR	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISGYO	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ISMEN	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
IZMDC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KARSN	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
KCHOL	0	0	3	0	0	0	0	0	0	5	4	0	4	0	0	0
KONYA	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
KORDS	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
KOZAA	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
KOZAL	0	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
KRDMD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAVI	0	0	0	0	0	0	0	0	0	3	3	0	3	0	0	0
MGROS	1	0	0	0	0	1	2	0	3	0	0	3	0	0	0	5
ODAS	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
OTKAR	2	0	0	1	0	0	1	0	1	0	0	1	0	0	0	1
OYAKC	1	0	0	2	0	1	0	0	2	0	0	2	0	0	0	2
PETKM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PGSUS	1	0	0	3	0	1	2	0	0	0	0	4	0	0	0	3

Appendix E 4 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
AEFES	3	0	0	4	0	0	0	0	0	0	3	0	0	0	0	3	0
AGHOL	5	0	0	3	0	0	0	0	0	0	3	0	0	0	0	3	0
AKBNK	3	0	0	4	0	0	0	0	0	0	3	0	0	0	0	3	0
AKCNS	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
AKFGY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKSEN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ALARK	5	0	0	3	0	0	0	0	0	0	3	0	0	0	0	3	0
ALBRK	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
ARCLK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASELS	0	5	1	0	0	3	0	3	2	0	0	0	0	0	2	0	0
ASUZU	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BAGFS	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
BERA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BIMAS	0	1	4	0	0	1	0	1	1	0	0	0	0	0	1	0	0
BRSAN	0	1	2	0	0	1	0	1	1	0	0	0	0	0	1	0	0
BRYAT	0	2	1	0	0	2	0	2	3	0	0	0	0	0	4	0	0
BUCIM	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
CCOLA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CEMTS	3	0	0	3	0	0	0	0	0	0	4	0	0	0	0	4	0
CIMSA	0	2	1	0	0	2	0	2	4	0	0	0	0	0	3	0	0
DOAS	3	0	0	3	0	0	0	0	0	0	4	0	0	0	0	4	0
DOHOL	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
ECILC	0	1	4	0	0	1	0	1	1	0	0	0	0	0	1	0	0
ECZYT	0	2	1	0	0	2	0	2	4	0	0	0	0	0	3	0	0
EGEEN	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
EKGYO	0	1	4	0	0	1	0	1	1	0	0	0	0	0	1	0	0
ENJSA	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
ENKAI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EREGL	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
FROTO	0	3	1	0	0	5	0	3	2	0	0	0	0	0	2	0	0
GARAN	3	0	0	4	0	0	0	0	0	0	3	0	0	0	0	3	0
GLYHO	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
GSDHO	3	0	0	3	0	0	0	0	0	0	4	0	0	0	0	4	0
GUBRF	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
HALKB	0	0	0	0	0	0	2	0	0	3	0	0	5	0	0	0	2
HEKTS	0	1	5	0	0	1	0	1	1	0	0	0	0	0	1	0	0
IPEKE	0	5	1	0	0	3	0	3	2	0	0	0	0	0	2	0	0
ISCTR	3	0	0	4	0	0	0	0	0	0	3	0	0	0	0	3	0
ISDMR	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
ISGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
ISMEN	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
IZMDC	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
KARSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KCHOL	3	0	0	3	0	0	0	0	0	0	5	0	0	0	0	5	0
KONYA	0	2	1	0	0	2	0	2	5	0	0	0	0	0	3	0	0
KORDS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KOZAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KOZAL	0	3	1	0	0	3	0	5	2	0	0	0	0	0	2	0	0
KRDMD	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
MAVI	3	0	0	4	0	0	0	0	0	0	3	0	0	0	0	3	0
MGROS	0	1	3	0	0	1	0	1	1	0	0	0	0	0	1	0	0
ODAS	0	0	0	0	0	0	2	0	0	2	0	0	2	0	0	0	4
OTKAR	0	2	1	0	0	2	0	2	4	0	0	0	0	0	3	0	0
OYAKC	0	1	2	0	0	1	0	1	1	0	0	0	0	0	1	0	0
PETKM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PGSUS	0	1	5	0	0	1	0	1	1	0	0	0	0	0	1	0	0

Appendix E 5 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	AEFES	AGHOL	AKBNK	AKCNS	AKFGY	AKSA	AKSEN	ALARK	ALBRK	ARCLK	ASELS	ASUZU	BAGFS	BERA	BIMAS	BRSAN
SAHOL	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
SASA	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
SELEC	0	0	0	1	0	0	0	0	0	0	1	0	0	0	5	2
SISE	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
SKBNK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	0	0	0	1	0	0	0	0	0	0	1	0	0	0	3	2
TAVHL	3	5	3	0	0	0	0	5	0	0	0	0	0	0	0	0
TCELL	0	0	0	2	0	0	0	0	0	0	5	0	0	0	1	1
THYAO	0	0	0	1	0	0	0	0	0	0	1	0	0	0	4	2
TKFEN	4	3	4	0	0	0	0	3	0	0	0	0	0	0	0	0
TKNSA	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0
TOASO	0	0	0	2	0	0	0	0	0	0	3	0	0	0	1	1
TSKB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TTKOM	0	0	0	2	0	0	0	0	0	0	3	0	0	0	1	1
TTRAK	0	0	0	5	0	0	0	0	0	0	2	0	0	0	1	1
TUKAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TUPRS	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESBE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESTL	0	0	0	3	0	0	0	0	0	0	2	0	0	0	1	1
YKBNK	3	3	3	0	0	0	0	3	0	0	0	0	0	0	0	0
ZOREN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix E 6 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	BRYAT	BUCIM	CCOLA	CEMTS	CIMSA	DOAS	DOHOL	ECILC	ECZYT	EGEEN	EKGYO	ENJSA	ENKAI	EREGL	FROTO	GARAN
SAHOL	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	3
SASA	0	0	0	5	0	5	0	0	0	0	0	0	0	0	0	3
SELEC	1	0	0	0	1	0	1	5	1	1	5	1	0	0	1	0
SISE	0	0	0	5	0	5	0	0	0	0	0	0	0	0	0	3
SKBNK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	1	0	0	0	1	0	1	3	1	1	3	1	0	0	1	0
TAVHL	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	3
TCELL	2	0	0	0	2	0	2	1	2	2	1	2	0	0	3	0
THYAO	1	0	0	0	1	0	1	4	1	1	4	1	0	0	1	0
TKFEN	0	0	0	3	0	3	0	0	0	0	0	0	0	0	0	4
TKNSA	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOASO	2	0	0	0	2	0	2	1	2	2	1	2	0	0	5	0
TSKB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TTKOM	2	0	0	0	2	0	2	1	2	2	1	2	0	0	3	0
TTRAK	3	0	0	0	4	0	5	1	4	5	1	5	0	0	2	0
TUKAS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TUPRS	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	3
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESBE	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
VESTL	4	0	0	0	3	0	3	1	3	3	1	3	0	0	2	0
YKBNK	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	3
ZOREN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix E 7 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	GLYHO	GSDHO	GUBRF	HALKB	HEKTS	IPEKE	ISCTR	ISDMR	ISGYO	ISMEN	IZMDC	KARSN	KCHOL	KONYA	KORDS	KOZAA
SAHOL	0	4	0	0	0	0	3	0	0	0	0	0	5	0	0	0
SASA	0	5	0	0	0	0	3	0	0	0	0	0	4	0	0	0
SELEC	0	0	1	0	4	1	0	0	0	1	0	0	0	1	0	0
SISE	0	5	0	0	0	0	3	0	0	0	0	0	4	0	0	0
SKBNK	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	0	0	1	0	3	1	0	0	0	1	0	0	0	1	0	0
TAVHL	0	3	0	0	0	0	3	0	0	0	0	0	3	0	0	0
TCELL	0	0	2	0	1	5	0	0	0	2	0	0	0	2	0	0
THYAO	0	0	1	0	5	1	0	0	0	1	0	0	0	1	0	0
TKFEN	0	3	0	0	0	0	4	0	0	0	0	0	3	0	0	0
TKNSA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOASO	0	0	2	0	1	3	0	0	0	2	0	0	0	2	0	0
TSKB	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
TTKOM	0	0	2	0	1	3	0	0	0	2	0	0	0	2	0	0
TTRAK	0	0	5	0	1	2	0	0	0	5	0	0	0	5	0	0
TUKAS	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
TUPRS	0	4	0	0	0	0	3	0	0	0	0	0	5	0	0	0
ULKER	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
VAKBN	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
VESBE	0	0	0	0	0	0	0	3	2	0	0	0	0	0	0	0
VESTL	0	0	3	0	1	2	0	0	0	3	0	0	0	3	0	0
YKBNK	0	4	0	0	0	0	3	0	0	0	0	0	5	0	0	0
ZOREN	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0

Appendix E 8 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	KOZAL	KRDMD	MAVI	MGROS	ODAS	OTKAR	OYAKC	PETKM	PGSUS	SAHOL	SASA	SELEC	SISE	SKBNK	SNGYO	SOKM
SAHOL	0	0	3	0	0	0	0	0	0	0	4	0	4	0	0	0
SASA	0	0	3	0	0	0	0	0	0	4	0	0	5	0	0	0
SELEC	1	0	0	3	0	1	2	0	4	0	0	0	0	0	0	3
SISE	0	0	3	0	0	0	0	0	0	4	5	0	0	0	0	0
SKBNK	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	1	0	0	5	0	1	2	0	3	0	0	3	0	0	0	0
TAVHL	0	0	3	0	0	0	0	0	0	3	3	0	3	0	0	0
TCELL	3	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
THYAO	1	0	0	3	0	1	2	0	5	0	0	4	0	0	0	3
TKFEN	0	0	4	0	0	0	0	0	0	3	3	0	3	0	0	0
TKNSA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOASO	3	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
TSKB	0	0	0	0	2	0	0	0	0	0	0	0	0	5	0	0
TTKOM	5	0	0	1	0	2	1	0	1	0	0	1	0	0	0	1
TTRAK	2	0	0	1	0	4	1	0	1	0	0	1	0	0	0	1
TUKAS	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0
TUPRS	0	0	3	0	0	0	0	0	0	5	4	0	4	0	0	0
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0
VESBE	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESTL	2	0	0	1	0	3	1	0	1	0	0	1	0	0	0	1
YKBNK	0	0	3	0	0	0	0	0	0	5	4	0	4	0	0	0
ZOREN	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	0

Appendix E 9 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

	TAVHL	TCELL	THYAO	TKFEN	TKNSA	TOASO	TSKB	TTKOM	TTRAK	TUKAS	TUPRS	ULKER	VAKBN	VESBE	VESTL	YKBNK	ZOREN
SAHOL	3	0	0	3	0	0	0	0	0	0	5	0	0	0	0	5	0
SASA	3	0	0	3	0	0	0	0	0	0	4	0	0	0	0	4	0
SELEC	0	1	4	0	0	1	0	1	1	0	0	0	0	0	1	0	0
SISE	3	0	0	3	0	0	0	0	0	0	4	0	0	0	0	4	0
SKBNK	0	0	0	0	0	0	5	0	0	2	0	0	2	0	0	0	2
SNGYO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SOKM	0	1	3	0	0	1	0	1	1	0	0	0	0	0	1	0	0
TAVHL	0	0	0	3	0	0	0	0	0	0	3	0	0	0	0	3	0
TCELL	0	0	1	0	0	3	0	3	2	0	0	0	0	0	2	0	0
THYAO	0	1	0	0	0	1	0	1	1	0	0	0	0	0	1	0	0
TKFEN	3	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3	0
TKNSA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOASO	0	3	1	0	0	0	0	3	2	0	0	0	0	0	2	0	0
TSKB	0	0	0	0	0	0	0	0	0	2	0	0	2	0	0	0	2
TTKOM	0	3	1	0	0	3	0	0	2	0	0	0	0	0	2	0	0
TTRAK	0	2	1	0	0	2	0	2	0	0	0	0	0	0	3	0	0
TUKAS	0	0	0	0	0	0	2	0	0	0	0	0	3	0	0	0	2
TUPRS	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	5	0
ULKER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VAKBN	0	0	0	0	0	0	2	0	0	3	0	0	0	0	0	0	2
VESBE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VESTL	0	2	1	0	0	2	0	2	3	0	0	0	0	0	0	0	0
YKBNK	3	0	0	3	0	0	0	0	0	0	5	0	0	0	0	0	0
ZOREN	0	0	0	0	0	0	2	0	0	2	0	0	2	0	0	0	0

Appendix E 10 Bilateral movement Alphabet Size = 4 Number of Segment = 25 Combination for Ward Linkage (Continued)

CURRICULUM VITAE

2016 – 2021	B.Sc., Industrial Engineering, Abdullah Gül University, Kayseri, TÜRKİYE
2021 – 2024	M.Sc., Abdullah Gül University, Kayseri, TÜRKİYE
2022 – Present	Present Research Assistant, Industrial Engineering, Abdullah Gül University, Kayseri, TÜRKİYE

SELECTED PUBLICATIONS AND PRESENTATIONS

J1) M. E. Nalici, İ. Soylemez, ve R. Ünlü, "Symbolic Aggregate Approximation-Based Clustering of Monthly Natural Gas Consumption", Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, c. 13, sy. 1, ss. 307–313, 2024, doi: 10.17798/bitlisfen.1395411.

J2) Nalici, M. E., & Akbaş, A. (2022). Forecasting of Occupancy Rate of Dams in İstanbul. Avrupa Bilim Ve Teknoloji Dergisi(41), 229-239. https://doi.org/10.31590/ejosat.1084484

C1) M. E. Nalici, İ. Söylemez, R. Ünlü Clustering of Climate Change Across Countries: Analysis of Surface Temperature and CO2 Emissions 12th Global Conference on Global Warming (May. 2024).

C2) M. E. Nalici, İ. Söylemez, R. Ünlü Symbolic aggregate approximation-based clustering of monthly natural gas consumption in 5th International Conference on Global Practice of Multidisciplinary Scientific Studies (Dec. 2023).