Jurnal E-Komtek
Vol. 7, No. 2 (2023) pp. 357-365
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

Optimization of Vertical Axis Darrieus Wind Turbine As Aerator Generator

Kiagus Rachmat Fauzan[1], Agus Pramono[2], Abdul Azis[3]
[1,2]Department of Information Systems, Universitas Amikom Purwokerto, Indonesia, 53127
[3]Department of Informatics, Universitas Amikom Purwokerto, Indonesia, 53217

fauzanrachmat@gmail.com

https://doi.org/10.37339/e-komtek.v7i2.1416

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
30-11-2023
Revised:
12-12-2023
Accepted:
12-12-2023
Online first :
12-12-2023

Abstract

The recent surge in data science or analytics in finance is crucial for informed decision-making in financial operations. Its integration is instrumental in helping businesses mitigate operational risks, identifying and preventing potential threats such as system failures and lapses. The study methodology involves case investigations and tests utilizing K-Means, K-Nearest Neighbors (KNN), and Decision Tree algorithms. Data visualization, using various graphical representations like graphs, diagrams, maps, and other techniques, is essential to simplify data interpretation and foster intuitive analysis. Common forms include heat maps, scatter plots, pie charts, bar graphs, line graphs, and maps, contributing significantly to presenting data comprehensively, offering insightful analyses, and achieving research objectives..

Keywords: Data Science, Data Visualization, Data Modeling, Financial, Stock Market

Abstrak

Lonjakan baru-baru ini dalam ilmu data atau analitik di bidang keuangan sangat penting untuk pengambilan keputusan yang tepat dalam operasi keuangan. Integrasinya sangat penting dalam membantu bisnis memitigasi risiko operasional, mengidentifikasi dan mencegah potensi ancaman seperti kegagalan sistem dan penyimpangan. Metodologi penelitian ini melibatkan investigasi kasus dan pengujian dengan menggunakan algoritma K-Means, K-Nearest Neighbors (KNN), dan Decision Tree. Visualisasi data, dengan menggunakan berbagai representasi grafis seperti grafik, diagram, peta, dan teknik lainnya, sangat penting untuk menyederhanakan interpretasi data dan mendorong analisis intuitif. Bentuk yang umum termasuk peta panas, plot sebaran, diagram lingkaran, grafik batang, grafik garis, dan peta, yang berkontribusi secara signifikan dalam menyajikan data secara komprehensif, menawarkan analisis yang mendalam, dan mencapai tujuan penelitian.

***Kata-kata kunci**: Data Sciece, Data Visualization, Data Modeling, Keuangan, Stock Pasar*

1. Introduction

The development of the Internet has been a significant factor in the growth of data. The Internet has fundamentally changed the computing and communications sector, providing an equitable infrastructure foundation for information provision, a channel for information dissemination, and a means for collaboration and interaction between individuals and their computers without geographical restrictions [1]. Data science has been growing in Indonesia, and various initiatives and programs are aimed at promoting data science education and application in the country [2]. Overall, the development of data science in Indonesia is still in its infancy. However, various initiatives and programs aim to promote the education and application of data science in the country. With the increasing need for data scientists and the increasing availability of data, the application of data science in Indonesia is expected to become more widespread [3]. Descriptive statistics is a method used to summarize and describe the main features of a dataset. It is commonly used in finance to analyze financial data, including investment and stock market data. Descriptive statistics provide a clear and concise summary of data, which can be used to identify patterns, trends, and relationships in the data [4]. One can use graphical and numerical techniques to perform a descriptive analysis of financial data. Graphical methods include using histograms, box plots, and scatter plots.

In contrast, numerical methods include using central statistical parameters such as mean, median, and mode and dispersion statistical parameters such as variance and standard deviation [5]. In the context of investment and stock market data, descriptive statistics can be used to analyze the performance of individual stocks or portfolios and compare the performance of different stocks or portfolios. For example, an investor may use descriptive statistics to analyze a particular stock or portfolio's historical returns to make informed investment decisions [6]. Descriptive statistics are useful tools for analyzing financial data, including investment and stock market data. It provides a clear and concise summary of data, which can be used to identify patterns, trends, and relationships in the data and to make informed investment decisions. Data science has various applications in the financial industry. Several steps need to be taken to apply data science in finance, including data extraction, validation, scrubbing, mining, and interpretation [7]. During the data scrubbing process, data is standardized and irrelevant or missing data is removed. During the data mining process, statistical analysis and data visualization are used to identify patterns and trends in the data. Finally, practical

communication skills are required to communicate the insights gained from the data to stakeholders during the data interpretation process **[8]**.

The application of data science or data analytics in finance is critical and has grown in recent years. There are many ways in which data science can be applied in finance, such as **[9]**:

a. Stock market analysis and investment: In this case, science data can predict stock market prices, assess investment risk, and identify the optimal investment portfolio to maximize profits.
b. Credit risk analysis: In banking activities, science data can be used to assess a customer's credit risk. This can help banks to issue loans with minimal risk and faster approval.
c. Credit card usage analysis: In this case, science data can analyze customers' credit card usage, purchase behaviour, and spending tendencies. This can help banks and credit card companies develop new marketing strategies and products more suitable for customers.
d. Fraud activities such as money laundering, insurance fraud, and so on. With data science, fraud patterns can be detected faster, and action can be taken to avoid losses.
e. Applying data science in finance is essential to help improve smart and accurate business decisions. Data science can help companies to manage operational risks. In this case, it can be used to analyze business operations, helping to identify and prevent system breakdowns, operational failures and other risks that could impact the business.

2. Method

This research applies the Descriptive Method in conducting Financial Data Analysis related to Investment and the Stock Market by considering the attributes of Gender and Age as research variables. The Descriptive Method is used to explain in detail the variables relevant to the problem under study. Within the framework of this method, the analysis is carried out through the following steps: Use Case Analysis, application of K-nearest neigh bours, Decision Tree (KNN), and K-means algorithms. Each method involves the process of data collection, algorithm modelling, implementation, visualization, and accuracy testing, which can be seen in **Figure 1 [3]**.

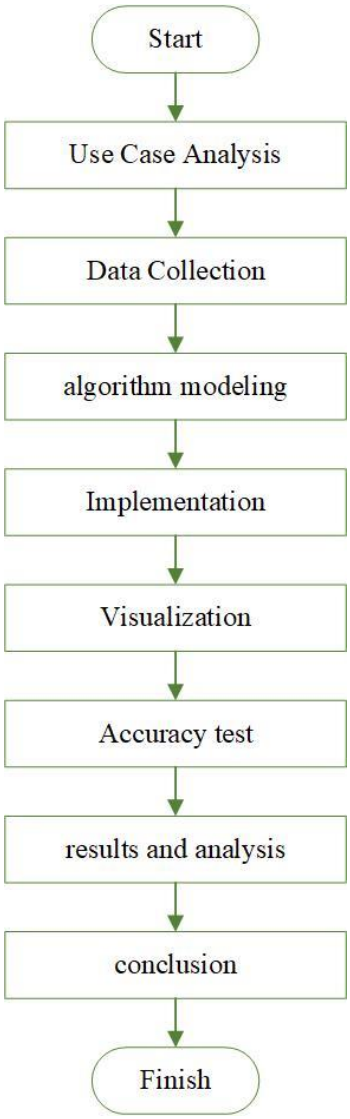

Figure 1. Research Method

The results of this analysis are then used to summarize the findings and derive relevant conclusions. The data used in the research came from a publicly accessible source, namely a CSV file titled "Finance_data.csv", downloaded from the kaggle.com website. The data explored in this study includes attributes of Gender and Age.

3. Results and Discussion

The analysis results used descriptive methods of financial data analysis on investment and stock markets based on gender and age attributes using data visualization. Data visualization is a process that aims to describe a pattern and information contained in a dataset visually through various graphical tools such as graphs, diagrams, maps, and other visualization methods **[10]**. This approach aims to improve data analysis and understanding more intuitively and understandably for researchers. Some common examples of types of data visualization

include line graphs, heatmaps, bar charts, pie charts, maps, and scatter plots. With the help of data visualization, complex information can be presented more clearly, facilitating better understanding and supporting more informed decision-making in a research context. Researchers used the Heatmap model in the Use Case Analysis to see the correlation between age and debentures.

3.1 Data Visualization using Use Case Analysis with Heatmap Model

The visualization presented above illustrates the correlation values between attributes, where the darkness of the colour on the value reflects the level of correlation between variables. The darker the colour of the value, the lower the level of correlation, which can be interpreted as a negative correlation. Conversely, the lighter the colour of the value indicates a positive correlation between the two variables. In this context, the visualization shows a positive correlation between the attributes of age and debentures, with a positive correlation value of 0.33. Data visualization using use case analysis with heatmap model is presented on **Figrue 2**.

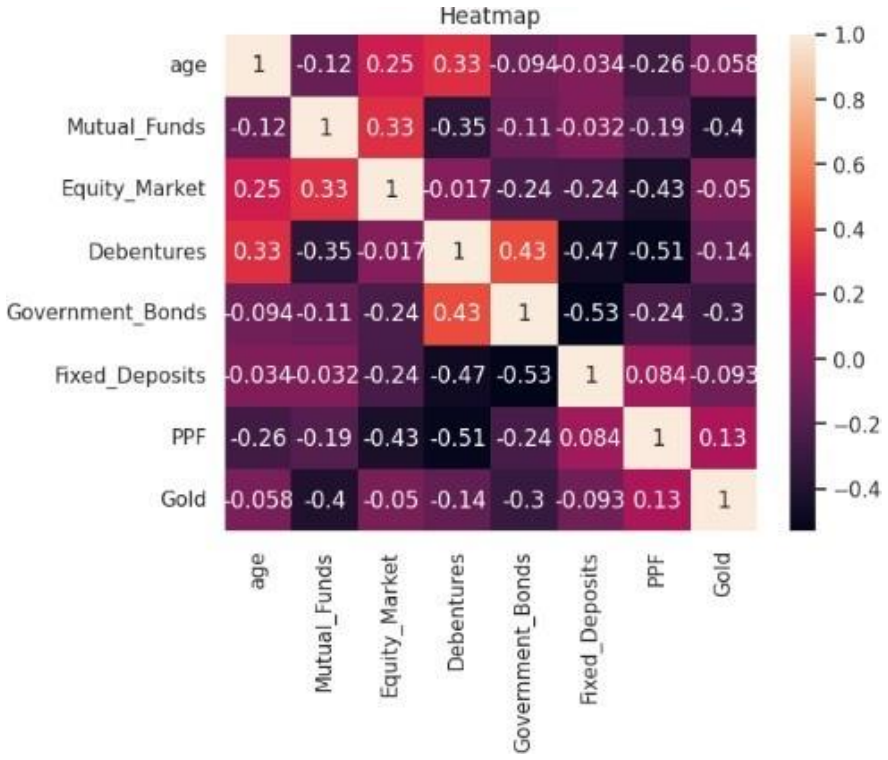

Figure 2. Data Visualization using Use Case Analysis with Heatmap Model

3.2 Data Visualization with KNN Method

I was testing the algorithm with the KNN method. The KNN method determined that the accuracy of the gender and age attributes was 87.5%; this value is relatively high and worth using in analyzing financial data. Here is the visualization generated from the KNN method. Data visualization with KNN method is presented on **Figure 3**.

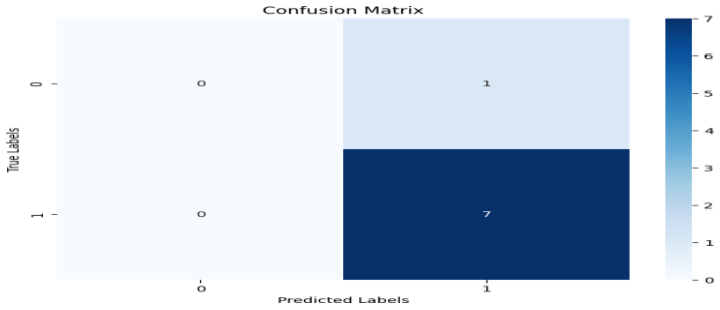

Figure 3. Data Visualization with KNN Method

3.3 Visualization of Clustering Results of the K-Means method

Testing by applying the K-Means method presents a visualization of the clustering results performed using the method. Holistically, a scatter plot is created in the presented set of codes that visualizes the data clustered using the K-Means method. In this plot, the data points are coloured according to the labels or classes the algorithm has generated. In contrast, the centroid points (cluster centres) are shown with different colours to distinguish them. The plot's axes and the accompanying title provide the context for this visualization to be well understood. Visualization of clustering results of the k-means method is presented on **Figure 4**.

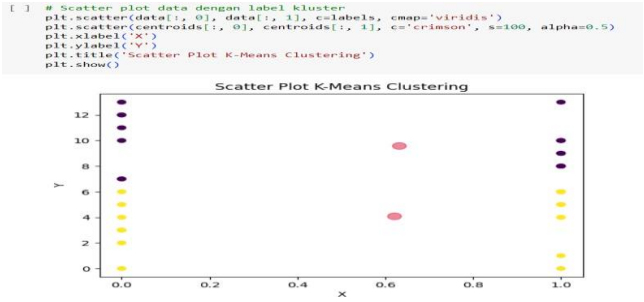

Figure 4. Visualization of Clustering Results of the K-Means Method

3.4 Visualization results of the Decision Tree Method

In the implementation of the Decision Tree method conducted by the researcher, an accuracy rate of 87.5% was achieved for the gender and age attributes. This achievement shows significant accuracy and relevance in analyzing financial data. The decision tree visualization in this study is a graphical representation of a tree structure consisting of nodes and branches connecting them. Each node in the decision tree represents a particular rule or classification in the dataset. The branches emanating from the node reflect various potential options or divisions based on the features present in the data. At each node, there is a certain amount of information that can be observed, including **[3]**:

a. Feature Usage: For each node, the variables used for the data separated at that stage are identified, for example, such as "gender" or "age."

b. Splitting criteria: The range of values used in splitting the data at that level is indicated, such as "age <= 30" or "gender = Female".
c. Number of samples: Information about the number of data samples included in the node.
d. The number of samples in each class: The distribution in the target class at the node is depicted.
e. The value of the majority class: The target class with the most samples at the node is displayed.
f. Predicted class: The node's target class indicates when a sample placed in the node is identified.

Visualization results of the decision tree method is presented on **Figure 5**.

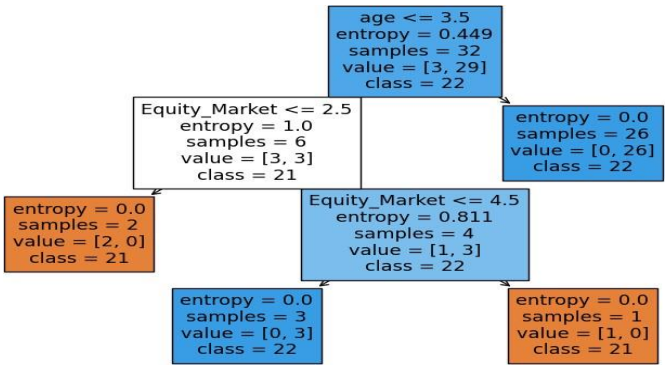

Figure 5. Visualization results of the Decision Tree method

By observing the Decision Tree visualization, the researcher can monitor the model's decision-making process based on the dataset's characteristics. In addition, the researcher can also understand how each data split is performed at each node in the tree structure and how the target class prediction is finally formed at the leaf nodes in the decision tree.

4. Conclusion

Based on the results of research that researchers have conducted and the previous discussion, it can be concluded that:

a. The importance of visualization in the context of datasets is evident, as it is an essential tool in illustrating data holistically. Visualizations provide sharper and deeper insights, which are key to achieving specific data analysis goals.
b. Based on the results of the K-Means model, it can be seen that the data accuracy achieved is low, to be precise at 50%. This means the model can only correctly cluster or classify about half of the data. This result indicates the potential for improvement or expansion of the model

to increase the ability to group more accurately. Scatter plot visualization of clusters can be used to understand the relationship between data points and identify outliers. Furthermore, K-Means' low accuracy can only sometimes be used to conclude that there is no correlation between the tested attributes. It is essential to consider other analysis methods and use appropriate statistical tools to answer the question of correlation between the tested attributes more accurately and validly.

c. Based on the KNN and Decision Tree models, the accuracy of the data is the same at 87.5% with different random state values in the training and test data.

References

[1] E. F. Prasandri, "Analisis Financial Distress Dengan Menggunakan Metode Z-Score (Altman), Springate, Dan Zmijewski Untuk Memprediksi Kebangkrutan Perusahaan Rokok Yang Terdaftar Di Bei Pada Tahun 2013-2016," *Jurnal Penelitian Ekonomi Dan Akuntansi*, Vol. 12, No. 2, Pp. 713–724, 2018.

[2] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Edik Informatika*, Vol. 7, No. 1, Pp. 213–219, 2023.

[3] R. Puspita And A. Widodo, "Perbandingan Metode Knn, Decision Tree, Dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan Bpjs," *Jurnal Informatika Universitas Pamulang*, Vol. 5, No. 4, Pp. 646–654, Dec. 2021, Doi: 10.32493/Informatika.V5i4.7622.

[4] H. Hiyanti, L. Nugroho, C. Sukmadilaga, And T. Fitrijanti, "Peluang Dan Tantangan Fintech (Financial Technology) Syariah Di Indonesia," *Jurnal Ilmiah Ekonomi Islam*, Vol. 5, No. 3, Pp. 1–8, Jan. 2020, Doi: 10.29040/Jiei.V5i3.578.

[5] D. Triyansyah And D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Incom Tech, Jurnal Telekomunikasi Dan Komputer*, Vol. 8, No. 3, Pp. 163–182, 2018, Doi: 10.22441/Incomtech.V8i2.4174.

[6] E. Budiono, "Analisis Financial Knowledge, Financial Attitude, Income, Locus Of Control, Financial Management Behavior Masyarakat Kota Kediri," *Jurnal Ilmu Manajemen (Jim)*, Vol. 8, No. 1, Pp. 284–295, 2020.

[7] A. K. Ngamel, "Analisis Finansial Usaha Budidaya Rumput Laut Dan Nilai Tambah Tepung Karaginan Di Kecamatan Kei Kecil, Kabupaten Maluku Tenggara," *Jurnal Sains Terapan Edisi Ii*, Vol. 2, No. 1, Pp. 39–47, 2012.

[8] E. Subiyantoro And F. Andreani, "Analisis Faktor-Faktor Yang Mempengaruhi Harga Saham (Kasus Perusahaan Jasa Perhotelan Yang Terdaftar Di Pasar Modal Indonesia)," *Jurnal Manajemen & Kewirausahaan*, Vol. 5, No. 2, Pp. 171–180, 2003, [Online].

[9] K. Ridha Pranita And F. Titik Kristanti, "Analisis Financial Distress Menggunakan Analisis Survival," *Nominal: Barometer Riset Akuntansi Dan Manajemen*, Vol. 9, No. 2, Pp. 240–257, 2020, [Online].

[10] A. W. G. P. Putro And T. Setiadi, "Penerapan Klasifikasi Decision Tree (C4.5) Untuk Memprediksi Kelulusan Siswa Sekolah Dasar Di Kecamatan Juai," *Jurnal Format*, Vol. 12, No. 2, Pp. 151–157, 2023.