

CHAPTER 1 INTRODUCTION

1.1. Background of the Problem

In the digital age and technological advancements, social media has become an important platform for users to share their thoughts, opinions, and feelings on a variety of topics, including sporting events such as the Asian Cup. Therefore, it is important to understand people's views and opinions by analyzing social media data.

The Asian Cup is a football tournament organized by the Asian Football Confederation (AFC). The Asian Cup was held every four years from 1956 to 2004 because the Summer Olympics and the European Football Championship were also held in the same year as the Asian Cup (2004, 2008, 2012, etc.). After 2004, the next Asian Cup was held in 2007 or three years later, and every four years (2011, 2015, 2019, etc.). The 2024 Asian Cup Qatar will host the 18th Asian Cup. This event officially takes place in winter and lasts for a month, from January 12 to February 10, 2024, with 24 participating teams.

As mentioned in Allah's words about *hifdz 'aql* (Reason to think) in sentiment analysis research on major events such as the Asian Cup:

إِنَّ فِي خَلْقِ السَّمُوتِ وَالْأَرْضِ وَاخْتِلَافِ اللَّيْلِ وَالنَّهَارِ لَآيَاتٍ لِّأُولِي الْأَلْبَابِ. الَّذِينَ يَذْكُرُونَ
اللَّهَ قِيَامًا وَقُعُودًا وَعَلَىٰ جُنُوبِهِمْ وَيَتَفَكَّرُونَ فِي خَلْقِ السَّمُوتِ وَالْأَرْضِ رَبَّنَا مَا خَلَقْتَ هَذَا بَاطِلًا
سُبْحَانَكَ فَقِنَا عَذَابَ النَّارِ

Meaning: Indeed, in the creation of the heavens and the earth, and the alternation of night and day, there are signs (of Allah's greatness) for the sensible person. (i.e.) those who remember God while standing, sitting, or lying down, and think about the creation of the heavens and the earth (saying, "O our Lord, thou did not create all this in vain. Most Holy Thy. Protect us from the punishment of hell. (Q.S. Al-Baqarah Verses 190-191).

The verse teaches the importance of deep and reflective thinking. In the field of sentiment analysis, maintaining reason means encouraging rational, data-based analysis, and free from bias or prejudice that damages intellectual integrity.

Sentiment analysis is included in the field of natural language processing (NLP) and is a process used to help determine the content of a data set in the form of opinions or viewpoints (sentiment) in written form about a positive, negative or neutral issue or event.^{1,2}.

In analyzing sentiment, the Naïve Bayes method has become a commonly used option. In addition, there are various other sentiment classification methods, such as Support Vector Machines, Random Forest, K-Nearest Neighbors, Neural Networks, and Logistic Regression that can provide more comprehensive insights.

Naive Bayes Classifier (NBC) is a probability-based classification method, where a set of probabilities is calculated by summing the frequency and combination of values from a given dataset³.

Each method of sentiment classification has advantages and disadvantages. Naïve Bayes is simple, fast and efficient, but relies on assumptions of independence that may not always be fulfilled. Other methods offer advantages in terms of skills in dealing with data complexity, although they may have their own challenges.

The Support Vector Machine (SVM) method has advantages in its ability to handle high-dimensional datasets and the complexity of relationships between features. Support Vector Machine is also effective in dealing with classification problems on data that is linearly inseparable and has good generalization capabilities. However, Support Vector Machine can require careful parameter adjustment and require significant computing resources especially on large datasets. Meanwhile, Random Forest offers the advantage of overcoming the problem of overfitting by building multiple decision trees and combining their results. The model is also capable of handling high-dimensional data and identifying important features in the analysis. However, its disadvantages include a lack of direct interpretability and the possibility of consuming large computing resources, especially in the formation of many trees. The choice between the Support Vector Machine and the Random Forest should consider the characteristics of the data and the specific analysis needs.

¹ Fanissa, S., Fauzi, A. M. and Adinugroho, S. (2018). "Analysis of Tourism Sentiment in Malang City Using the Naive Bayes Method and Selection of Query Expansion Ranking Feature | Journal of Information Technology Development, SMATIKA Journal Volume 10 Number 02, December 2020 ISSN: 2087-0256, e-ISSN: 2580-6939 and Computer Science," J. Information Technology and Computer Science Development., vol. 2, no. 8, pp. 2766– 2770.

² Chandani, V., Computer, I. F. and Nuswantoro, D. U. (2015). "Comparison of Machine Learning Classification Algorithms and Feature Selection on Film Review Sentiment Analysis," J. Intell. Syst., vol. 1, no. 1, pp. 56–60.

³ AlHaq, T., & Arfian, A. (2021). Twitter user sentiment analysis on Covid-19 vaccination using the Naive Bayes method. *Inti Nusa Mandiri* 2021, 15(2), 1–8.

Based on the background that has been explained, in this study, sentiment analysis will be carried out on netizens' comments on social media X towards the 2024 Asian Cup in Qatar in 2024 using the Naïve Bayes algorithm. The purpose of this study is to find out the sentiment and results of the accurate classification of social media user sentiment X about the Asian Cup in Qatar in 2024 using Naïve Bayes.

1.2. Problem Formulation

Based on the problems described in the background above, the formulation of the problem in this study is the interaction of the Indonesian people on the pros and cons of the overall match aspect, the performance of the match equipment, and naturalized players at the 2024 Qatar Asian Cup on social media platform X, as well as the extent to which the Naïve Bayes method can provide an accurate analysis of these sentiments.

1.3. Problem Limitations

For the formulation of a specific problem, the limitations of the problem in this study are as follows:

- The Programming Language in this study is Python
- This system classifies the views of the Indonesian people on the overall aspect of the match, the performance of the match equipment, and naturalized players in the Asian Cup in Qatar 2024 with a neutral, positive, or negative classification using only the Naïve Bayes method.
- Data collection on this system only uses social media platform X, using the keywords "Naturalization", "AFC referee", and "AsianCup2023". Starting from November 1, 2023, to March 1, 2024.
- The data that was successfully collected was 2050 data, with the division into 3 categories, namely "Naturalization" with 783 data, "AFC Referees" with 683 data, and "AsianCup2023" with 584 data.

1.4. Research Objectives

From the formulation of the problem that has been written above, the purpose of the research is to classify the sentiment of the Indonesian people towards the overall match aspect, the performance of the match equipment, and naturalized players at the 2024 Qatar Asian Cup using data obtained from X. With the Naïve Bayes method in neutral, positive or negative form.

1.5. Research Benefits

The benefits that the author expects from this study are as follows:

1. For Students

This research is useful as an insight enhancer and delves deeper into sentiment analysis, especially in the naïve bayes method.

2. For Other Researchers

The usefulness of this study in other researchers is as a reference for future research with better research with the evaluation of previous research.

3. For Readers

This research will be knowledge for readers about the biggest competition in Asia, namely the Asian Cup and also the countries of powerful countries and great players who participate in the competition.

1.6. Systematics of Discussion

This research uses the following writing systematics:

CHAPTER 1: INTRODUCTION

1.1. Background of the Problem

1.2. Problem Formulation

1.3. Problem Limitations

1.4. Research Objectives

1.5. Research Benefits

1.6. Systematics of Discussion

CHAPTER 2: LITERATURE REVIEW

2.1. Previous Research

2.2. Theoretical Foundations

CHAPTER 3: RESEARCH METHODOLOGY

3.1. Time and Place of Research

3.2. Research Tools and Materials

3.3. Research Stages

CHAPTER 4: RESULTS AND DISCUSSION

4.1. Results

4.2. Results of the Naive Bayes Method

4.3. Results of Netizen Sentiment Distribution

4.4. Deployment

CHAPTER 5: CONCLUSION

5.1. Conclusion

5.2. Suggestion

BIBLIOGRAPHY

APPENDIX – APPENDIX

