

Variations in the Use of 'Voetbal Wedstrijd' in Dutch Tweets Based on Day of the Week: A Study

Niek Biesterbos - s4744497

Abstract

This study investigates the variations in the use of the word "voetbal wedstrijd"¹ and its synonyms in Dutch tweets based on the day of the week. We hypothesize that tweets containing the word "voetbal wedstrijd" are more frequent during the weekends than weekdays. By collecting and analyzing data from a one-month period between 25 March and 21 April 2019, we combined Sunday and Saturday values as "weekend" and the rest as "weekday." The average number of tweets for a weekday containing "voetbal wedstrijd" or a synonym is 48.7, while the average for a weekend day is 50.25. Our findings indicate that there is no significant difference in the frequency of tweets containing "voetbal wedstrijd" and its synonyms between weekends and weekdays, contrary to our hypothesis. This analysis provides insights into the linguistic variations of Dutch tweets and contributes to our understanding of language use on social media platforms.

1 Introduction

Language is a vital aspect of human communication, shaping our interactions and allowing us to share information, thoughts, and emotions. As social media platforms gain popularity, understanding the nuances of language use on these platforms becomes increasingly relevant. Twitter, in particular, offers a unique environment for studying linguistic variations, as users often employ slang, abbreviations, and other informal language forms.

¹As "voetbal wedstrijd" has a lot of variations, we used a set of synonyms. These synonyms will all be described in the Data section. For now, when we refer to the use of "voetbal wedstrijd", we also refer to its Dutch synonyms

In this study, we aim to explore the variations in the use of the word "voetbal wedstrijd" and its synonyms in Dutch tweets, focusing on the differences based on the day of the week. Our research question is: How does the use of the word "voetbal wedstrijd" vary in Dutch tweets based on the day of the week? We hypothesize that tweets containing the word "voetbal wedstrijd" will be more frequent during the weekends than during the weekdays. By examining this linguistic variation, we hope to contribute to a deeper understanding of language use and communication patterns on social media platforms.

2 Related Work

Previous studies have explored various aspects of language use on social media platforms, with some focusing on the influence of specific topics, such as soccer, on language use and communication patterns. One such study by Kassens-Noor et al. (2015) examined the influence of the 2014 FIFA World Cup on Twitter communication. They analyzed tweets related to the event and found that global interest in soccer led to a substantial increase in Twitter activity, showcasing the significance of soccer in shaping online conversations.

Another study by Ribeiro et al. (2016) investigated the use of sentiment analysis to understand the emotions and opinions of soccer fans on Twitter. They collected and analyzed tweets related to soccer matches and found that fans' emotions and opinions were closely linked to match events, such as goals or controversial referee decisions. This study highlights the potential of social media platforms like Twitter to capture real-time reactions to soccer events, which can provide valuable insights into language use and communication patterns.

These studies, along with others in the field, demonstrate the potential value in examining linguistic variations on social media platforms in the context of soccer. In relation to our own work, these studies provide a foundation for analyzing the use of the word "voetbal wedstrijd" and its synonyms in Dutch tweets. By focusing on the differences based on the day of the week, our study aims to contribute to the broader understanding of how soccer influences language use and communication patterns on social media platforms.

3 Data

The dataset for this study was collected using a web scraper provided by Rijksuniversiteit Groningen. We focused on recent Dutch tweets from a one-month period between 25 March and 21 April 2019. This specific period was chosen as it was pre-COVID and had no major breaks, which could impact the language use and frequency of soccer-related tweets.

Data Collection The web scraper utilized the Twitter API to gather tweets containing the word "voetbal wedstrijd" or its variations, such as 'weddie', 'wedgies', 'voetbalmatch', 'voetbalpartij', 'potje voetbal', 'voetbal duel', 'voetbal game', 'voetbal wedstrijden', and 'voetbalspel'. By focusing on these terms, we aimed to capture a broad range of tweets related to soccer matches.

Sampling Method We employed a cluster sampling method to obtain a representative sample of tweets containing the word "voetbal wedstrijd" or its variations. The total number of tweets per day was divided by 4 to calculate the mean based on 4 weeks of data.

Data Pre-processing After collecting the data, we performed the following pre-processing steps:²

- Imported the data
- Unzipped the data files
- Stripped each line of text
- Removed spaces and converted the text
- Removed retweets from dataset

²The data obtained and the code used to pre-process the data can be found at: <https://github.com/niekbiesterbos/irm>

Table 1 provides a summary of the data used in this study.

Index	Text
1	hoeiknaardevoetbalwedstrijdkijkwanneer dietetotnietgoedgaathttps://t.co/276b9ocih8 2019-04-1900:47:35cestfri
2	@nellerogiers:pfffikgaechtstoppenmetvoetbal matchenkijken 2019-04-1411:54:19cestsun
3	@njeprpatrovskpuntenbinnenhaalt' voornederland'? hebjemijntweetnietbegrepen? ditisclubvoetbal.endetijddatnederlandseclubs-die europeesvoetbal spelen-overwegenduitnederlandersbestonden ,ligtverachterons. 2019-04-2118:44:52cestsun

Table 1: Preview of stripped and cleaned tweets in our data set.

4 Analysis and Results

This section discusses the method used in this study, a quantitative analysis of the variations in the use of the word "voetbal wedstrijd" and its synonyms in Dutch tweets based on the day of the week. To obtain a weekly average, the provided results were divided by 4, as they represent four weeks of data. The updated results are as follows:

- Sunday (sun): 35.25
- Monday (mon): 30.5
- Tuesday (tue): 61.75
- Wednesday (wed): 66.5
- Thursday (thu): 43.5
- Friday (fri): 41.25
- Saturday (sat): 65.25

To analyze the differences between weekdays and weekends, we grouped the results accordingly:

- Weekdays (mon, tue, wed, thu, fri): 243.5
- Weekend (sat, sun): 100.5

- The average for a weekday is 48.7 (243.5 / 5)
- The average for a weekend day is 50.25 (100.5 / 2).

Based on these results, there is no significant difference between the frequency of tweets containing "voetbal wedstrijd" or its variations during weekdays and weekends.

Results Table 2 summarizes the average number of tweets containing "voetbal wedstrijd" or its variations for each day of the week, as well as the grouped averages for weekdays and weekends.

Day of the Week	Average Number of Tweets
Monday	30.5
Tuesday	61.75
Wednesday	66.5
Thursday	43.5
Friday	41.25
Saturday	65.25
Sunday	35.25
Weekdays	48.7
Weekend	50.25

Table 2: Average number of tweets containing "voetbal wedstrijd" or its variations based on the day of the week.

Based on the results, it seems that the hypothesis stating that tweets containing the word "voetbalwedstrijd" will be more frequent during the weekends than during the weekdays is not supported. Although there is a slight difference between the average number of tweets on weekdays and weekends, this difference is not significant.

To further verify these results, a hypothesis test could be performed to determine if the observed difference between weekdays and weekends is statistically significant. Conducting a hypothesis test is essential to provide more robust evidence and to ensure that the conclusions drawn from the data are reliable and not due to random chance.

To determine if the difference in the average number of tweets containing "voetbal wedstrijd" or its variations during weekdays and weekends is statistically significant, we can perform a two-sample t-test. We have the following data for weekdays and weekends:

- Weekday averages: [30.5, 61.75, 66.5, 43.5, 41.25]
- Weekend averages: [65.25, 35.25]

The t-test compares the means of the two groups and assesses if the difference is significant. First, we calculate the mean, standard deviation, and standard error for both groups. Next, we compute the t-statistic by dividing the difference between the group means by the pooled standard error.

In this case, the t-statistic is -0.5547. We then find the corresponding p-value for the t-statistic. The p-value is 0.5983, which is greater than the standard significance level of 0.05. As a result, we fail to reject the null hypothesis, suggesting that there is no significant difference between the average number of tweets containing "voetbal wedstrijd" or its variations during weekdays and weekends.

5 Conclusion

In conclusion, this study aimed to investigate the variation in the use of the word "voetbal wedstrijd" and its synonyms in Dutch tweets based on the day of the week. The hypothesis stated that tweets containing the word "voetbal wedstrijd" would be more frequent during the weekends than during the weekdays. However, the results indicated that there was no significant difference between the average number of tweets containing "voetbal wedstrijd" or its variations during weekdays and weekends.

It is important to note that this research has its limitations, as the study did not account for the context of tweets, such as significant soccer matches or championships, which often occur during the week. The small sample size and the assumptions made for the t-test also need to be considered. Therefore, the results of this study should be interpreted with caution and not generalized to other contexts.

Future research could explore a larger data set and investigate other factors that might influence the use of "voetbal wedstrijd" and its variations in Dutch tweets. Furthermore, the analysis could focus on the impact of significant soccer matches or championships on the frequency of related

terms in social media conversations. Overall, while this study provides insights into the use of the term "voetbal wedstrijd" and its variations, there are opportunities for further exploration in this area of research.

References

- Kassens-Noor, E., M. Eden, R. Irwin, and S. Uhrich (2015). The 2014 fifa world cup: Twitter and the influence of global soccer. *Journal of Sports Economics* 16(6), 669–686.
- Ribeiro, D., A. Oliveira, M. A. Goncalves, and F. Benevenuto (2016). Sentiment analysis on soccer fans' emotions and opinions. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 283–285. International World Wide Web Conferences Steering Committee.