Problem Chosen	2025 MCM/ICM Summary Sheet	Team Control Number
ABCDEF		12345678

The MCM Thesis of Team 12345678

Summary

This is a summary.

Keywords: keyword1, keyword2, keyword3

Contents

1 Introduction

1.1 Problem Background

Wordle, developed by Jonathan Feinberg in 2008, was created to help students expand their vocabulary. However, due to its simple gameplay, it quickly went viral on social media at the end of 2021 and was later acquired by The New York Times in 2022, integrating it into their online games section. It is a web-based game with two difficulty modes: easy and hard. It focuses on user experience and game logic, and there are many variations of the game, such as Quordle (guessing 4 words simultaneously), Octordle (guessing 8 words simultaneously), and Worldle (a geography version where players guess a country or region). The rules for the hard mode are as follows.

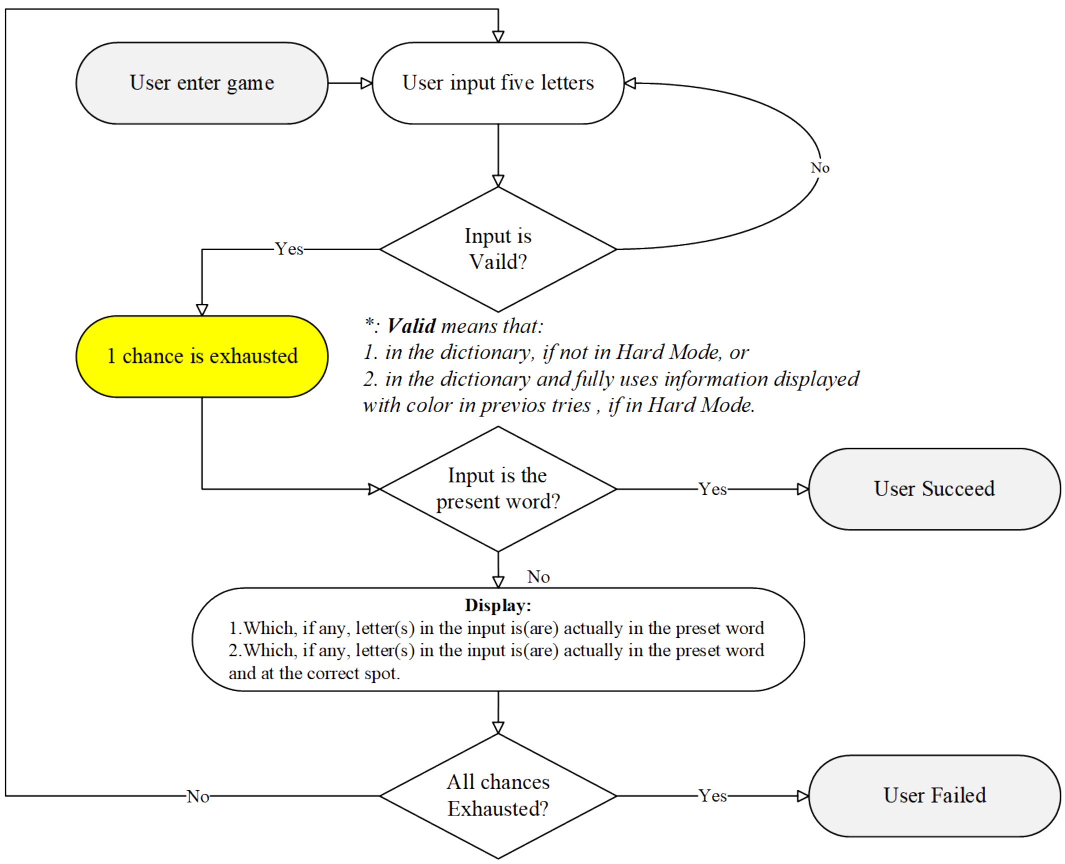

Figure 1: Game Rules

1.2 Restatement of the Problem

We need to analyze the data provided by The New York Times and address the following tasks:

- **Problem 1:** Develop a model to explain the variations in the daily reported results and predict the range of reported results for March 1, 2023. Additionally, analyze which word attributes influence players' decisions to select Hard Mode.

- **Problem 2:** Build a prediction model to estimate the percentage distribution of results (1, 2, 3, 4, 5, 6, X) for a future day, with specific predictions for "EERIE" on March 1, 2023, and assess the model's accuracy.
- **Problem 3:** Develop a classification model to categorize words by difficulty level and identify their attributes. Conduct a detailed analysis for "EERIE" and evaluate the model's accuracy.
- **Problem 4:** Explore and describe any other interesting insights or patterns found within the data.

1.3 Our Work

2 Assumptions and Notations

2.1 Assumptions

To simplify the model, we made the following assumptions:

- **Assumption 1:** The percentages provided in the data accurately reflect the actual situation with minimal falsification and can be used directly.
  Justification: Most people report their scores truthfully, as the reporting process is anonymous and has no consequences. Although some individuals may choose to report false scores, the daily volume of reports is large enough that this small fraction of false data will not significantly impact the overall percentages.
- **Assumption 2:** The percentages of total reported scores can be used to represent the percentages of scores reported in Hard Mode.
  Justification: Due to rounding, we found that the sum of the daily reported percentages does not equal exactly 1 but still reaches 0.99, which is very close to 1. Therefore, the reported percentages can effectively substitute for the actual percentages.

2.2 Notations

Symbol	Decription
f_t	forget gate
i_t	input gate
o_t	output gate
h_t	hidden state
c_t	cell state
x_t	LSTM's input
$W_{f,i,c,o}$	Bias
$b_{f,i,c,o}$	Weight Matrix
r	Pearson correlation

3 Model 1-Prediction Model based on LSTM

By analyzing $Problem\_C\_Data\_Wordle.xlsx$, we found that the Wordle data is collected and analyzed daily. By analyzing this data, we can clearly see that the results of Wordle change over time and exhibit time dependence. Therefore, we chose the LSTM algorithm, which is specifically designed to capture and utilize long-term dependencies in sequential data. This fits well with the time-varying Wordle results presented in the problem. We ultimately used this algorithm to simulate the data for March 1, 2023.

3.1 Description of LSTM

LSTM is a neural network structure that can effectively capture long-term dependencies in time series data. By introducing gating mechanisms, it can selectively retain and forget information, solving the vanishing and exploding gradient problems inherent in raditional RNNs. Although LSTM is computationally expensive, it performs excellently in many tasks, especially in processing sequential data. Its main features are the memory cell and gating mechanisms, which introduce three gates: the forget gate, input gate, and output gate, allowing for selective retention or discarding of data. Its update equations are as follows.

$$
\begin{aligned}
&f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
&i_t \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
&o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
&\tilde{c}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\
&c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\
&h_t = o_t \cdot tanh(c_t)
\end{aligned}
\tag{1}
$$

3.2 Prediction on March 1,2023

3.2.1 Data settings

Machine learning algorithms are very sensitive to the input data values. To prevent certain results (e.g., Christmas 2022) from being significantly smaller or larger than other features, which could cause biases in the algorithm's processing of data, we applied a formula to temporarily scale all data to the range [0, 1] to facilitate computation. The formula is as follows.

$$
x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{2}
$$

We then divided the given data into training and test sets. After analyzing the data, we found that the results of Wordle generally showed a decreasing trend. When the proportion of the training set was not particularly large (7:3 or even 8:2), due to the lack of early and late results, the final predicted result would be much higher or lower than the actual data from December 31, 2022. Therefore, we chose a 9:1 data split. We also decided to use data from the previous month (30 days) to predict the results. It is well known that during statutory holidays, the statistics may significantly drop because people need to spend time with family and friends. According to website data[1], in high-GDP countries, there are about 12 statutory holidays a year on average, which means that there is almost always one holiday per month. Therefore, using the previous 30 days of data would almost cover a holiday, leading to results that better reflect the actual situation. Finally, we used the formula below for inverse normalization to restore the data to its original form.

$$x_{original} = x_{scaled} \times (max(x) - min(x)) + min(x) \tag{3}$$

3.2.2 Results on March 1,2023

After setting the parameters, we ran the model. Since the data for March 1 is about 60 days after December 31, we first used data from October 1 to December 31 (about 90 days) for preliminary fitting to observe the model's prediction results. The specific data is shown in Figures 2.

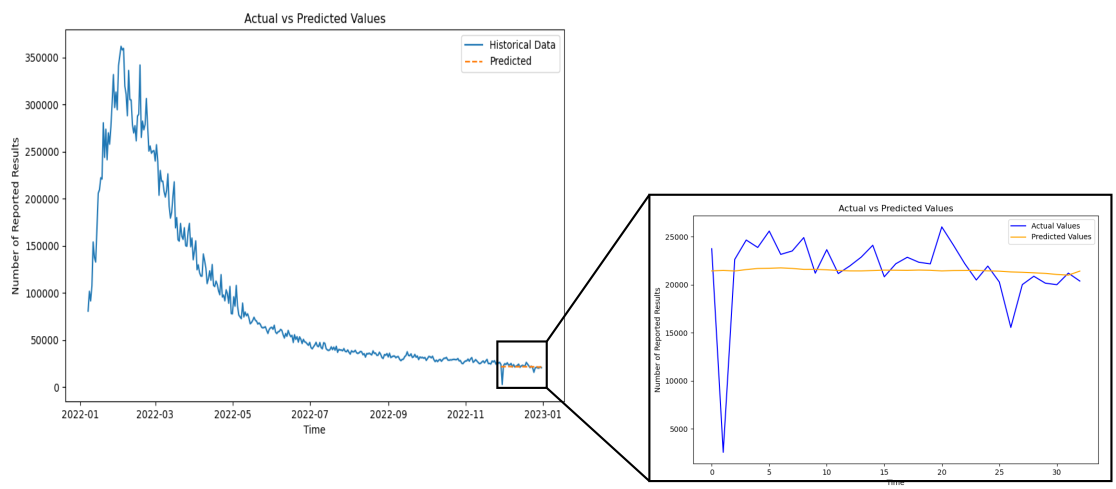

Figure 2: Actual vs Predicted Values

We found that the results of this fitting had a good correlation with the actual values, so we also used the model to predict the next 90 days. The results are shown in Figure 3.

We also tracked the changes in training loss and validation loss over time. As the number of iterations increased, the training loss gradually decreased, indicating that our model was increasingly fitting the training data and learning the underlying patterns. Eventually, it stabilized around 2% at the 10th iteration, suggesting that the model had

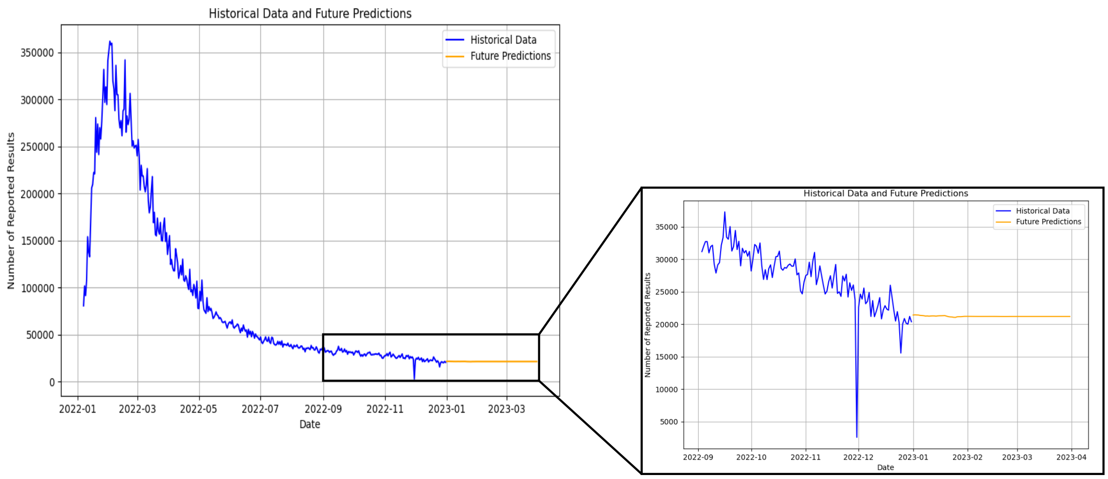

Figure 3: Historical Data and Future Predictions

converged and reached an optimal point. To prevent overfitting, we also closely monitored the changes in validation loss. Initially, we observed that the validation loss increased during the early iterations, likely due to the model adapting to the training data. However, after four iterations, it started to decrease, showing signs of generalization. By the 10th iteration, the validation loss had also stabilized, indicating that the model was no longer overfitting. We took the error value at the 10th iteration as the final result, and the data plot is shown in Figure 4, reflecting the training dynamics throughout the process.

Based on our model, the predicted range for the number of reported results on March 1, 2023, is:

$$x_{March1,2023} = 21137 \pm 2.01425\% \tag{4}$$

4 Relationship of Word Attributes and Scores from Percentage

4.1 Score setting

The data provided in the problem is presented in percentages, which does not accurately reflect the actual scores for a given day. Therefore, we define the score for a specific day as follows:

$$score = w_1 * \sum_{i=1}^{6} i * p_i + w_2 * p_x \tag{5}$$

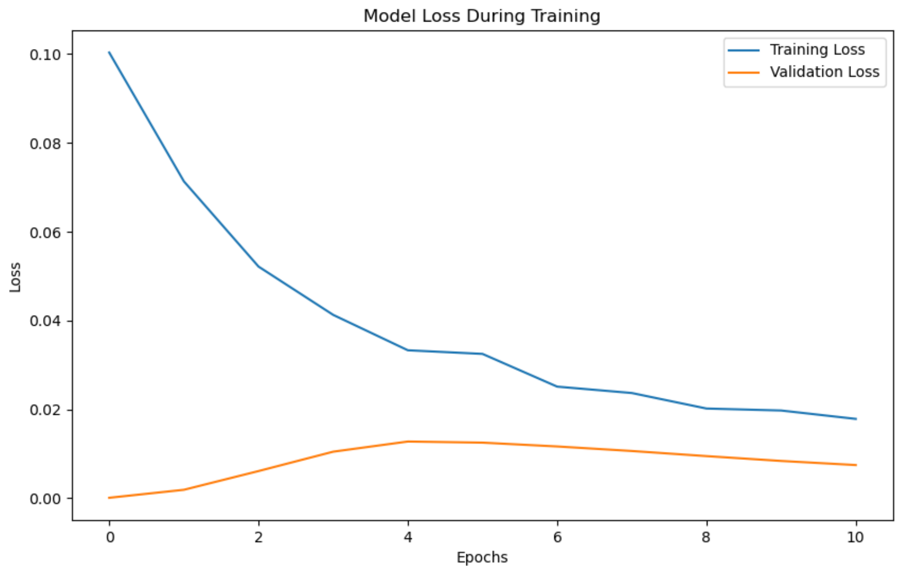

Figure 4: Model Loss During Training

The score is determined by the number of correct guesses, meaning that 4 successful attempts would result in a score of 4. Since the game only has two possible outcomes, success or failure, we assign a weight of 0.5 to each outcome in order to balance the results.

4.2 Frequency setting

Considering that people from various industries are playing Wordle, we began to think about which types of words are more likely to be guessed. Since people from different industries have varying levels of expertise, their vocabulary also tends to differ, but the words they commonly use are generally those used in everyday life, such as "sleep," "lunch," and so on. Therefore, we believe that the frequency of a word's appearance will impact the results. To address this, we used google website[2] to collect statistics on the frequency of each word used in Wordle puzzles, denoted as f_{word}.

4.3 Regression Analysis

After defining frequency and score, we set them as the x-axis and y-axis, respectively, and plotted a scatter plot in Figure 5. Additionally, we created heatmaps in Figure 6 for frequency, score, and the number of attempts, and calculated the Pearson correlation to measure their correlation, The formula is as follows:

$$r = \frac{\sum(f_{word,i} - \bar{f_{word}})(score_i - sc\bar{o}re)}{\sqrt{\sum(f_{word,i} - \bar{f_{word}})^2}\sqrt{\sum(score_i - sc\bar{o}re)^2}} \tag{6}$$

From the plots, we observed that the correlation between a word's frequency and its

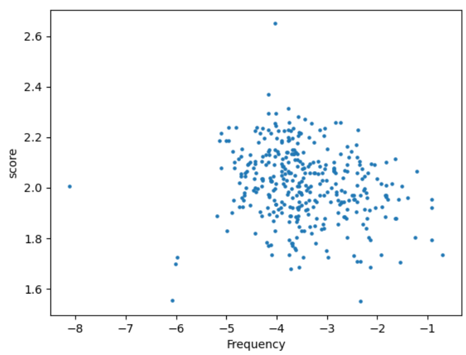

Figure 5: Word frequency

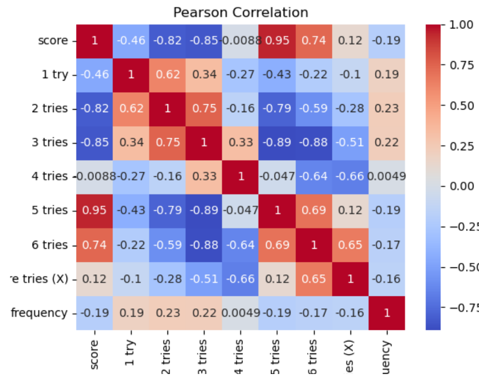

Figure 6: Pearson Correlation

score is only 0.19, indicating a weak correlation. We believe this is because, even though some words are frequently used in daily life, they may have synonyms that are more commonly used, causing the original words to be forgotten. For example, the word "mummy" from October 23, 2022. Additionally, some words have had their original meanings replaced by other associations. For instance, the word "watch" from March 11, 2022—due to the prevalence of smartphones, people now check the time on their phones rather than looking at a watch.Ultimately, we concluded that the frequency of a word's usage is u nrelated to its performance in difficult modes.

5 Model 2

6 Model 3

7 Interesting Findings

7.1 4 Tries is the watershed

What we are interested in is that through the heatmap of frequency and score, we found that 4 tries have almost no correlation with score and 5 tries, but there is a certain correlation with 3 tries. Therefore, we believe this is because the majority of players are able to arrive at the correct answer within 3 to 4 attempts, which keeps 4 tries in a dynamic balance. This makes 4 tries act as a constant in the statistical score, while a portion of players can only guess the correct answer in 5 tries or 6 tries. These two groups of players do not interfere with each other, which is why there is no correlation. We specifically plotted the correlation between 4 tries and 5 tries, and 3 tries for easier viewing. The detailed image is shown below.

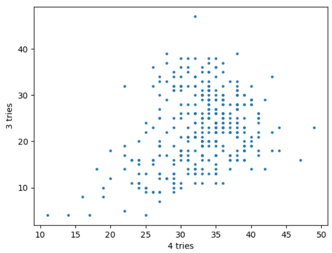

Figure 7: 4 and 3 tries'

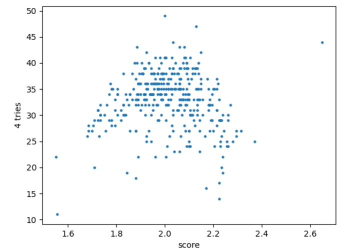

Figure 8: 4 tries and score's correlation

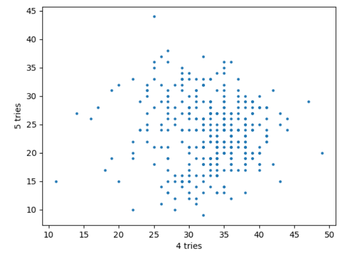

Figure 9: 4 and 5 tries'

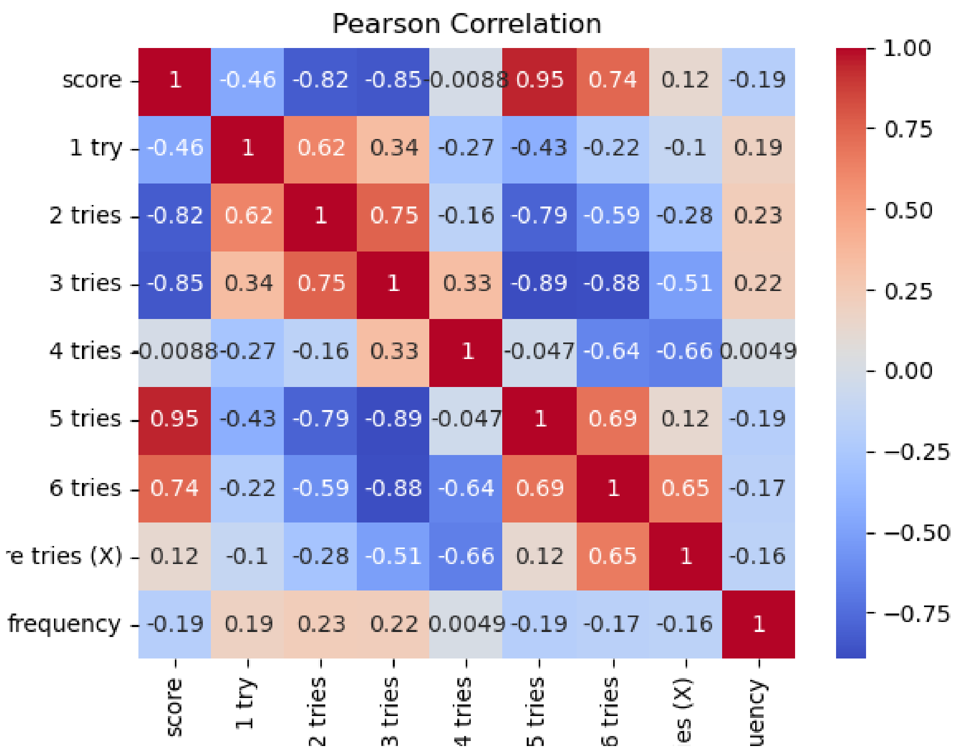

Figure 10: Pearson Correlation

8 Sensitivety Analysis

9 Model Assessment

9.1 Strengths

9.2 Weaknesses

10 Letter

I love math.

I love math.

I love math.

References

[1] https://holidays-calendar.net/

[2] https://books.google.com/ngrams/

Appendices

MEMORANDUM

To: MCM office

From: MCM Team 12345678

Subject: MCM

Date: January 15, 2025

This is a memorandum.

Appendix A First appendix

Here are simulation programmes we used in our model as follow.
MATLAB source code:

Appendix B Second appendix

Python source code: