When the epidemic was raging, Wordle became a popular game on the Internet due to its extremely low operation complexity, high game fun and unique sharing mechanism. In order to help the New York Times carry out better development planning, game optimization and management, we respectively established the ARIMA model, BPNN model and K-means algorithm base model to help the New York Times carry out data prediction and analysis.

As for Problem 1, we built two models. In order to predict the number interval of future report results, we established the ARIMA model by analyzing the overall fluctuation trend of the data given for 359 days. The data of the first 328 days were used as the training set of the model to predict the data of the next 31 days. After comparing the predicted results with the real data, it was found that there was a very high degree of coincidence between them. We used all 359 data as the training set and predicted that the number of reported results of 2023/3/1 was within the range of 16735~17238. In order to find the relationship between word attributes and the ratio of the number of hard mode reports to the total number of reports, we took three factors as parameters: letter frequency, combined sound frequency and word frequency (processed by sigmoid function) of words, and compared their occurrence frequency in hard mode respectively. Finally, it is found that there is no direct correlation between the word attributes and the ratio of difficulty patterns.

As for Problem 2, we use BP neural network model to predict the distribution of reported data. On the basis of known input parameter 7 and output result 7, the predicted value and the real value under different overfitting times and different hidden layers were compared, and the accuracy of the model was evaluated with RMSE as the measurement standard. The results show that the prediction results are better when the number of overfitting is 100 and the number of hidden layers is 10. The model is used to predict the data of the word "EERIE" (Table 5), and the distribution of its report results is shown in Table 6.

As for Problem 3, The K-mean model was used to divide the difficulty of words into four categories and evaluate them with seven parameters. For visualization purposes, the seven parameters were converted into two parameters, which are displayed in the rectangular coordinate system (Figure 13). For the example of "eerie", the correlation analysis is shown in Figure 14

Some other interesting observations we make during the meta-analysis of the data are that the number of results reported skyrocketed to a peak in the first 25 days of the data set and then declined daily; there was an abrupt low in the December 25 results report; the proportion of results of choosing difficult mode keeps increasing.

At last, we analyze the strengths and weaknesses of our model as well as its sensitivity, whose results show that our model has high robustness, precision and accuracy. After that, a letter to the Puzzle Editor of the New York Times is attached.