Record of project progress and plans

|  |  |  |  |
| --- | --- | --- | --- |
| Plan | How to do | Success criteria | deadline |
| On the Sentiment Analysis System: Data preprocessing | Cleaning of data, including removal of nulls, duplicates; word splitting and de-duplication of text. | The dataset is free of nulls and duplicates.  The text data is accurately segmented and deactivated. | 18/July |
| Feature engineering | The text data is vectorized using the TF-IDF method to generate the feature matrix. | The TF-IDF feature matrix is successfully generated, and the feature matrix can effectively represent the text content. | 21/July |
| Initial model training | Initial training was performed using an LSTM model with a TF-IDF feature matrix as input and sentiment scores as output. | A preliminary LSTM model was successfully trained, and the accuracy of the model on the training set exceeded 70%. | 27/July |
| Model evaluation | Model performance is evaluated using metrics such as accuracy, recall, and score. | 1.Obtain a full evaluation report.  2.Model has a score of 70% or more on the validation set. | 29/July |
| Time division experiment | Splitting the dataset by date, training and testing on data from different time periods, and evaluating the ability of the model to generalize over time. | 1.Successfully completed the time segmentation experiment.  2.The model performs consistently on data from different time periods, with scores varying by no more than 10 per cent | 1/August |
|  |  |  |  |
| The "usefulness score" model: initial model training | Initial training was carried out using the Random Forest model with a TF-IDF feature matrix as input and a usefulness score as output. | 1.Initial Random Forest model was successfully trained.  2.The accuracy of the model on the training set is more than 70%. | 8/August |
| Model evaluation | Model performance is evaluated using metrics such as accuracy, recall, and score. | 1.Obtain a full evaluation report.  2.The model has a score of 70 per cent or more on the validation set. | 10/August |
| Model adjustment | Based on the evaluation results, the model parameters are adjusted, and the model is tested. | 1.Successfully adjust the model and test the model.  2.The tuned model scores more than the initial model on the validation set. | 14/August |
|  |  |  |  |
|  |  |  |  |