# 5 Professional Issues

# 5.1 Project Management

## 5.1.1 Activities

All the activities related to the objectives are shown in the table below.

| Objects | Activity | Completed |
| --- | --- | --- |
| Ob1: Conduct background research on text classification, understand the field and the corresponding technologies. | A1.1 Identify subject keywords | Completed |
| A1.2 Search for relevant essays | Completed |
| A1.3 Read the relevant literature | Completed |
| A1.4 Summary the advantages and limitations of different technologies | Completed |
| A1.5 Perform a literature review | Completed |
| Ob2: Collect usable dataset from the Internet. | A2.1 Search for social media comment datasets on Kaggle | Completed |
| A2.2 Download the datasets | Completed |
| A2.3 Identify the structure of the datasets | Completed |
| Ob3: Clean and pre-process the data for modeling. | A3.1 Search for methods to clean the data | Completed |
| A3.2 Apply methods on datasets | Completed |
| A3.3 Evaluate the process | Completed |
| Ob4: Extract features from the text in the cleaned datasets. | A4.1 Search for methods to extract data | Completed |
| A4.2 Implement methods on the datasets | Completed |
| A4.3 Evaluate the feature extracting methods | Completed |
| Ob5: Train different models using datasets and assessing the quality of the models | A5.1 Search for documentation of different models | Completed |
| A5.2 Apply models in Python language | Completed |
| A5.3 Adjust the parameters until the model performs optimally | Completed |
| Ob6: Analyze the quality of the models and compare the strengths and weaknesses of each model | A6.1 Search for different methods to evaluate the model | Completed |
| A6.2 Apply multiple rubrics to different models | Completed |
| A6.3 Put the results into a table | Completed |
| Ob7: Develop data and model testing and evaluation strategy | A7.1Set up test plan for the dataset and model | Completed |
| A7.2 Put the test plan on the dataset and model | Completed |
| A7.3 Evaluate how well did the data and model perform | Completed |
| Ob8: Risk analysis based on current progress | A8.1 Search for risk analysis for deep learning projects | Completed |
| A8.2 Modify the risk analysis based on the used models | Completed |
| A8.3 Implement the analysis on the models | Completed |

Table 1 Activities of Objects

## 5.1.2 Schedule

The time schedule with the accomplished work and future work is shown below. The green color represents the accomplished work.

图表, 日程表

描述已自动生成

Figure 18 Schedule of the project period

## 5.1.3 Project Data Management

Whenever a new version of the code is updated, it will be uploaded to the Baidu Cloud in order to keep track of all project progress.

Whenever a new version of the report of the project or related electronic documentation is updated, it will be uploaded to the Feishu Shared Space in order to keep track of all project progress.

## 5.1.4 Project Deliverables

Throughout the execution of this project, the following items are submitted for assessment:

* Project proposal with ethical forms, showing detailed description of the work to be done. (Submitted)
* Project weekly report containing planned objectives for each week. (Submitted)
* Progress report providing justification of the project. (Submitted)
* Project presentation illustrated by a poster and a practical demonstration. (Submitted)
* Final report which comprises a complete and clear explanation of the problem to be solved. (Submitted)

## 5.2 Risk Analysis

Possible risks that may appear through the process of accomplishing the project with possible mitigation measures are listed below.

The table is arranged as:

Potential Risk: the possible risks that may appear in the process of the project

Potential Causes: The reason of having the risk

Severity: The impact degree potential causes may influence the project

Likelihood: The probability of the situation happening

Risk: The score of the Potential Causes

Mitigation: The method to prevent risk from happening

表格

描述已自动生成

Table 2 Risk Analysis

The scoring criteria for risks in the table are as follows. The likelihood of occurrence of the risk is divided into six levels, with the most likely risk score being 6, and the probability of occurrence decreasing in descending order. The severity of the risk is divided into four levels, with the worse the impact of the risk, the higher the score. The score of risk in the table is the product of likelihood and severity. Risk score below 6 is considered Acceptable (ACC), a score between 7 and 12 is evaluated as Require Further Investigation (RFI), while a score beyond 12 is defined as Not Acceptable (N/ACC).

## 5.3 Professional Issues

### 5.3.1 Legal Issue

This project strictly follows the regulations outlined in the General Data Protection Regulation (GDPR) [17] and is used to ensure that the data used for training and testing is legitimate, particularly with regard to transparency and purpose limitation.

### 5.3.2 Ethical Issue

The ethical issue raised by this project is that when the project is applied to real social media platforms, the accuracy of the model cannot be correct as many of the corpus cannot be updated in a timely manner due to the proliferation of new online words in today's online society. The misjudgment of non-malicious comments that results from this situation can lead to problems such as the banning of accounts of non-offending users, which may cause some legal concerns.

### 5.3.3 Social Issue

Most deep learning models are black box models. Because of the lack of intrinsic explanations of black box models, and the inevitably use of unbalanced sample data for model training, the use of such models can lead to a range of social problems such as gender bias and racial bias when the problem is severe.

### 5.3.4 Environment Issue

The main purpose of this model is to save human and time resources in social media platforms by automatically screening malicious comments by machines, which is environmentally friendly and greatly reduces the consumption of resources.