School of Computing

UNIVERSITY OF LEEDS

COMP5200M Project Specification

***NOTE to student**: ensure you have discussed the content with the supervisor. Submit an **electronic version** of this form in pdf via the COMP5200M module page on Minerva; with filename of the format <surname><year>-Spec (e.g. SMITH17-Spec.pdf).*

Student Name: Long Yue
Programme of Study: MSc Advanced Computer Science (Data Analytics)
Supervisor Name: Sebastian Ordyniak
Name of External Company (if any)**:**
Type of Project: Exploratory Software
Provisional Title of Project: Spam filtering model based on machine learning
Aim of Project: We receive a lot of emails in our daily study and work. In addition to emails related to study and work, we also receive a lot of spam, including advertising emails, fraudulent emails, and so on. This task model judges whether the email is normal (ham) or spam email (spam) according to the text content contained in the email, to realize automatic spam filtering.

Objectives:

By using open source data: Enron Email Dataset to design and train a spam filtering model based machine learning logic and methods. Hope to make and optimize a more accurate and fast filtering model to realize convenient mailbox management functions for individuals and companies. After the new model training is completed, it will be compared with the filtering models used in the past, and the model will be further optimized through comparison and evaluation, and the principles of machine learning used in this project will be explored and discussed.

Deliverables:

An accurate, reproducible and reuse filtering model. And machine learning filtering models are studied and discussed through evaluation and comparative studies. The model through data processing and massive training will have reliable mail filtering accuracy, and it is expected that in the future, the interface and application port can be further designed and imported into the mail application to realize the mail filtering function for users. And a report discussing in-depth model design, training, comparison and implementation process.