Minutes 2

|  |  |
| --- | --- |
| Subject: | COMP90051 Statistical Machine Learning |
| Group name/ identifier | Group 14 |
| Meeting Location, Date & Time | ERC Library, 07/09/2023 3.00pm |
| Group members present | Yinuo Sun, Yufeng Xie, Xueke Zheng |
| Apologies |  |

# Actions from previous meetings

|  |  |  |
| --- | --- | --- |
| **Task** | **Who is responsible** | **Deadline** |
| EDA of the raw data | Yufeng | 04/09 |
| Vectorization of the text | Xueke | 05/09 |
| Baseline model training | Yinuo | 07/09 |
|  |  |  |
|  |  |  |

# Items discussed and decisions made

|  |  |
| --- | --- |
| **Item** | **Decision (if applicable)** |
| Discuss about resampling methods: after EDA, we found that the text lengths of domain 1 & 2 are very different, hence multiple resampling methods or training strategy need to be tried | Try new methods of sampling/ separating domains |
| Baseline model improvement | Hyperparameter tuning |
| Discuss about more preprocessing/ feature engineering steps | Create a new feature of text length that can be used in model training |

# Agreed actions from this meeting

|  |  |  |
| --- | --- | --- |
| **Task** | **Who is responsible** | **Deadline** |
| Further preprocessing and feature engineering steps | All | 10/09 |
| Hyperparameter tuning and more resampling trials | Yinuo | 14/09 |
| Try deep learning models, e.g. CNN, RNN | Xueke, Yufeng | 15/09 |
|  |  |  |
|  |  |  |

# Next meeting

|  |  |
| --- | --- |
| Location | ERC Library |
| Date & Time | 15/09/2023 |
| Goals | Discuss about solution of any issues coming in the working process;  Think about any ways to improve model performances;  Solve difficulties in deep learning models. |