Meeting minutes 21/04/2022

|  |  |
| --- | --- |
| Subject: | COMP90051 Statistical Machine Learning |
| Group name/ identifier | Group53 |
| Meeting Location, Date & Time | Zoom, 21st April 2022, 21:30 |
| Group members present | * Yingdi Guan * Kiki Lan * Benjamin Zhou |
| Apologies |  |

# Actions from previous meetings

|  |  |  |
| --- | --- | --- |
| **Task** | **Who is responsible** | **Deadline** |
| Test out Naïve Bayes on the dataset | Kiki Lan | This meeting |
| Continue testing SVM to see if it works | Benjamin Zhou | This meeting |
| Come up with other ways to represent the data | Everyone | This meeting |
| Research if we can predict probabilities with KNN and if it is a good model for this task | Everyone | This meeting |
| Research other models we can potentially use | Everyone | This meeting |
| Start on section 1 and 3 of the report | Benjamin Zhou | This meeting |

# Items discussed and decisions made

|  |  |
| --- | --- |
| **Item** | **Decision (if applicable)** |
| What ML models will we use for the project? | * SVM: takes too long even with reduced dimensionality * Naïve Bayes: promising for an initial baseline model but seems maybe too simple for final model * KNN: too many training instances might result in too much overhead * Neural networks: still deciding |
| How should we represent the data (currently using a binary representation of keywords and coauthors)? | * Assign authors to research groups based on other authors they often write together with and ignore keywords (0.66787 on Kaggle submission) * Reduce the number of features to represent authors/keywords to the maximum number of authors/keywords found for any single paper in all training instances (119 features) |
|  |  |
|  |  |
|  |  |

# Agreed actions from this meeting

|  |  |  |
| --- | --- | --- |
| **Task** | **Who is responsible** | **Deadline** |
| Try to improve Naïve Bayes | Kiki Lan, Benjamin Zhou | Next meeting |
| Come up with other ways to represent the data | Everyone | Next meeting |
| Research other models we can potentially use | Everyone | Next meeting |
| Continue section 3 of report; include details about the new data representations tried | Benjamin Zhou | Next meeting |

# Next meeting

|  |  |
| --- | --- |
| Location | Zoom |
| Date & Time | Saturday 21:30 |
| Goals | Try out more models and data representations |