



CS985: Machine Learning for Data Analytics

Assignment Details

Part One - Relevance Classification Task.

Your notebook/report should contain each of the following sections and subsections in the order below.

- **Overview:** Provide a short summary (100-200 words)
 - Short summary of what you did, and your key message, learnings and recommendations given your modelling. (4-5 sentences).
- **Method:** Describe your methodology for modelling (300-400 words)
 - What did you learn from exploring the data that will influence how you approach modelling this dataset?
 - How did you process/pre-process the data (note it may be different for different models)
 - What is your test/train/validation strategy that you will be used.
 - I.e how will you split your data/classes/sampling/etc.
 - You will want to keep this consistent between models.
- **Models** (300-400 words - provide a short summary of what models/parameter tried)
 - Standard ML Baseline:
 - Create a simple baseline of your choice using standard ML
 - Hint. You will want to pre-process the text to transform it to be used by the ML, see: https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html
 - 3Layer NN Baseline:
 - Create a baseline 3 layer NN model with a feed-forward architecture
 - Vary the number of neurons in your single layer.
 - Show a graph of how adding more neurons changes performance
 - Hint. You will want to start off by exploring what happens, then pick a range to experiment and select 5-10 points to plot.
 - Where is the sweet spot? What is a good what to decide how many to use?
 - Deep NN models:
 - Extend your baseline to be a deep NN model, and find a good configuration (you should vary the number of layers, and choice of activation functions).
 - Show a line graph of how adding more layers Changes performance (try 3 - 5 combinations for your plot). Is deeper, better?
 - Show a bar plot of how different activations functions affect the performance. For this task what activation function works best?
 - Complex NN Models:
 - Rather than employing a standard deep NN model, this is your chance to use more sophisticated architectures. For example, RNNs, e.g., LSTM, GRUs, Wide and Deep Models, Bert, and other Large Language Models, to improve performance.
 - Try to find interesting architectures that might help improve the performance.
 - If you have 4 team members, submit one different model*.
 - If you have 5 team member, submit two different models
 - Experiment with the additional models, and explore the different parameters/configurations.
 - Include at least one graph per model showing how the performance changes as you change an interesting parameter of the model.
- **Results** (approx. 500 words):
 - Present a table of results showing your training/testing scores for each of the models that your team tried above.
 - Inc. the mean and standard deviations.
 - The table show report all the different configurations that you tried.
 - Provide a short summary of the findings given the performance you obtained, and what you tried, and what you learnt.
 - Explain what you tried, what worked, and what didn't.
- **Summary** (approx. 200 words)
 - Final Model and Performance
 - Which model do you recommend for this task? And why?
 - Your should submit your models to Kaggle — how well did it perform?
 - What next? What you you do next? How could you get better wins for this task?
- **References**
 - Include relevant references for models, techniques, etc that you have used.

Part Two - Emotion Recognition Classification Task

Your notebook/report should contain each of the following sections and subsections in the order below.

- **Overview:** Provide a short summary (100-200 words)
 - Short summary of what you did, and your key message, learnings and recommendations given your modelling. (4-5 sentences).

- **Method:** Describe your methodology for modelling (300-400 words)
 - What did you learn from exploring the data that will influence how you approach modelling this dataset?
 - How did you process/pre-process the data (note it may be different for different models)
 - What is your test/train/validation strategy that you will be used.
 - I.e how will you split your data/classes/sampling/etc.
 - You will want to keep this consistent between models.
 - Data Augmentation
 - You don't have that much data, employ different data augmentation strategies, to help your training.
- **Models** (300-400 words - provide a short summary of what models/parameter tried)
 - Standard ML Baseline:
 - Create a simple baseline of your choice using standard ML
 - Deep NN models:
 - Create a deep NN, and find a good configuration (explore at least 5-10 configurations)
 - Complex NN Models:
 - Rather than employing a standard deep NN model, this is your chance to use more sophisticated architectures. For example, CNNs, ResNet, etc. to improve performance.
 - Try to find interesting architectures that might help improve the performance.
 - If you have 4 team members, submit one different model*
 - If you have 5 team member, submit two different models
 - They must be different architectures (not just extra layers).
 - Experiment with the additional models, and explore the different parameters/configurations.
- **Results** (approx. 400 words):
 - Present a table of results showing your training/testing scores for each of the models that your team tried above.
 - Inc. the mean and standard deviations
 - Include the results with and without your different training schedules (i.e. with/without different data augmentation)
 - Provide a short summary of the findings given the performance you obtained, and what you tried, and what you learnt.
 - Explain what you tried, what worked, and what didn't.
- **Summary** (approx. 200 words)
 - Final Model and Performance
 - Which model do you recommend for this task? And why?
 - Your should submit your models to Kaggle — how well did it perform?
 - What next? What you you do next? How could you get better wins for this task?
- **References**
 - Include relevant references for models, techniques, etc that you have used.

Notes for Smaller Teams

- **If you have 4 team members, you only need to complete the complex models for part one or part two.**
- **If you have 3 teams member, you only need complete part one or part two of the assignment, but need to submit three complex models.**

Last modified: Monday, 18 March 2024, 7:50 PM

◀ Recommended source code for BERT

Jump to...

Assignment Cover Sheet ►

© University of Strathclyde

You are logged in as **Ayesha Michael Gaikwad (Log out)**

CS985

Home

Find a class

Student Experience

Registration

Visas & Immigration

Letters (Bank/Council Tax/Status)

Maintain Personal Details

Maintain Curriculum

Transfers & Exchanges

Exams

Personal Circumstances

Use of Social Media & Virtual Learning Environments

[Student Support & Wellbeing](#)
[Report & Support](#)
[Absence and Voluntary Suspension](#)
[Downloads for Students](#)
[Appeals](#)
[Complaints Procedure](#)
[Withdrawing from your studies](#)
[Graduation](#)
[Postgraduate Taught & Postgraduate Research courses](#)
[Transcripts](#)
[Academic Verification](#)
[Replacement Parchments](#)
[Replacement Student Card](#)
[Student Voice](#)
[Module Evaluations](#)
[Student Surveys](#)
[Report and Support](#)
[Module Evaluation Reports](#)
[You said, we did](#)
[StrathReps](#)
[Help](#)
[Services status](#)
[Myplace support](#)
[Request Support](#)
[Twitter](#)
[Create a site](#)
[Reset my password](#)