COMP90051 Assignment 1

- Deadline Apr 25, 2024 8:00 PM
- Deadline within the team
    - Code done/ debugged Apr 14, 2024
    - Report draft Apr 18, 2024
    - Final deadline Apr 21, 2024
- Canva Group 29

Overall Status

Actionables	Status	Notes	Related files
Assignment 1 Brief	N/A		COM…
Group Agreement	Launched		Grou…
Meeting Minute Template	Not star…	>= 3	Meet…

Meetings

Meeting 1 Apr 3, 2024 6:30 PM

Data processing

- Feature selection
    - Latent Semantic Analysis
    - PCA
    - LDA
- Oversampling
- Undersampling
- Tf-idf
- BoW
- word2vec

Possible approaches:

- Regression
- Model shift
- RNN
- LSTM

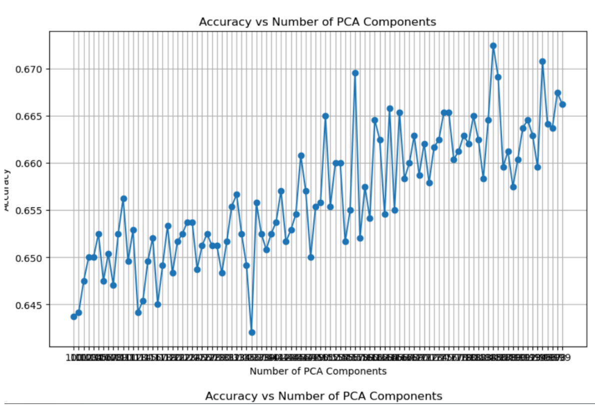
Problems to tackle
- Shrink the size of each word vector
- Find a way to train the model with significantly imbalanced data
- Handle potentially very different 2 domains

Meeting 2 Apr 6, 2024 7:00 PM

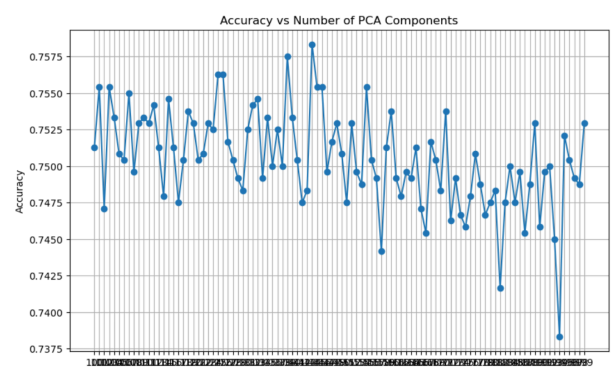
Agenda:
- Discuss which data processing and algorithm we would like to use
- Start coding~~
- Preprocessing
  - Bag of words
  - word2vec
  - Latent Semantic Analysis Chang Xu
  - PCA Daniel Manzano Aguayo
- Model
  - SVM - baseline
  - Neural network
  - RNN (Recurrent Neural Network)

Meeting 3 Apr 9, 2024 6:00 PM

- LSA Chang Xu
  - Did not improve the accuracy of both LR & SVM models
  - In the process of performing hyperparameters tuning
- PCA Daniel Manzano Aguayo
  - Hyperparameters tuning
  - Tried 100 - 150 models hasn’t really improved



Logistic regression



SVM

  - Plan to try tunning other parameters to improve accuracy
  - Combined it with LR & SVM
- Try a new method domain adaptation neural network Jijiang WEN

- domain shifting
- Up and down sampling the domain 2 data
    - Improve accuracy to 0.78
- Next step
    - Create more fake data – data augmentation
    - Try to implement domain shifting
    - Look for other alternatives to handle data that comes from different domains
        - Multi tasks
        - Ensembling learning
    - Hyperparameters tuning for LSA and PCA
    - Word2vec and bow
    - Tf idf(best)

Unsupervised Domain Adaptation by Backpropagation

Meeting 4 Apr 11, 2024 6:20 PM

- Ensembling learning
    - https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/
    - Adaboost Chang Xu
    - Combine Neural Networks with Adaboost
    - Stacking Daniel Manzano Aguayo https://www.analyticsvidhya.com/blog/2021/08/ensemble-stacking-for-machine-learning-and-deep-learning/
        - Random forest & logistic regression

- Svm lr rf knn -> combined domain 1 and 2 Jijiang WEN
- Bagging -> 4 models -> meta model Jijiang WEN

- Stop testing/using PCA or LSA
    - Might use it again in the future if takes too long to train the more complex models

Meeting 5 Apr 14, 2024 6:00 PM Chang Xu Daniel Manzano Aguayo

- Daniel Manzano Aguayo
- Approach: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

-
    - Train using all domain 1 & 2 data each domain with a different model.
        - Base models : SVC, SVC
        - Meta model: MLP
        - Downsampling domain 1 at the same size of domain 2.
        - Accuracy: 0.84 validation
        - Test: 0.65
    - Train using domain 1 & 2 together with different models
        - Base Models: LogisticRegression, KNNeighbours, RandomForest, Naive Bayes
        - Meta Model: MLP
        - Accuracy: 0.7956 validation
        - Test: 0.79650
- Chang Xu
    - Trained an adaboost and chose the hyperparameters using grid search
        - The best Kaggle result is 0.69900
    - Training a stacking model with adaboost and SVM
    - Will try https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/
    -

```
# combine the two datasets
combined_data = pd.concat([domain1_train_data,
domain2_train_data_balanced])


# get the features and labels
X = combined_data['text']
y = combined_data['label']


# split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# split the training data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train,
test_size=0.2, random_state=42)
```

Just split the dataset into training and validation.

Meeting 6 Apr 16, 2024 6:00 PM

- Discuss which model we should continue refining to try reach 0.85 or above accuracy
    - Rf model
    - Meta Model: MLP (second one)
-
    - Tf-idf -> standscaler(balanced combined data domain1 and domain2)
    - Tf-idf (balanced combined data domain1 and domain2) -> rf (para:)
    - 1500 1500
    - 5000 7000 weight oversampling (SMOTE)
        -

    Data augmentation

  Rnn LSTM

- Start the report?

  Intro

Meeting 7 Apr 18, 2024 6:00 PM

Online server for training model

- RNN 0.668 (accuracy with approach using tf-idf)
- Try RNN with word sequence approach

Transformer - jijiang wen and start report

Rf model:`Best parameters: {'max_depth': None, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 400}`

`Acc:0.82`

`data augmentation`

`3 words`

Meeting 8 Apr 21, 2024 6:00 PM

PTMDA

	text	label	id
0	[16, 231, 543, 5, 15, 43, 8282, 94, 231, 1129,...	1	0
1	[16, 4046, 138, 10, 2, 1809, 2007, 3763, 14, 4...	1	1
2	[1108, 16550, 3, 6168, 3, 160, 284, 19, 49, 46...	1	2

```
3  [1802, 27, 16, 25, 48, 451, 632, 3, 2, 2164, 2...      1   3
4  [16, 19, 302, 93, 97, 43, 952, 118, 1, 16, 528...      1   4
                                        text  label    id
0  [12, 920, 7, 1266, 28, 9884, 1640, 116, 11, 13...      1  5000
1  [783, 397, 253, 5797, 9379, 22, 793, 11838, 10...      1  5001
2  [888, 14851, 323, 9, 27, 1377, 584, 195, 3, 13...      1  5002
3  [228, 1161, 5815, 379, 9, 941, 10, 2, 316, 4, ...      1  5003
4  [736, 19, 37, 813, 45, 6723, 27, 626, 8, 2, 34...      1  5004

Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 150, 100)          1000000

 lstm (LSTM)                 (None, 150, 64)           42240

 dropout (Dropout)           (None, 150, 64)           0

 lstm_1 (LSTM)               (None, 32)                12416

 dense (Dense)               (None, 24)                792

 dense_1 (Dense)             (None, 1)                 25

=================================================================
Total params: 1,055,473
Trainable params: 1,055,473
Non-trainable params: 0
_________________________________________________________________
Epoch 1/10
200/200 [==============================] - 107s 518ms/step - loss: 0.6304 - accuracy: 0.6520 - val_loss: 0.6054 - val_accuracy: 0.6687
Epoch 2/10
200/200 [==============================] - 89s 446ms/step - loss: 0.5202 - accuracy: 0.7522 - val_loss: 0.5709 - val_accuracy: 0.7262
Epoch 3/10
200/200 [==============================] - 68s 341ms/step - loss: 0.4226 - accuracy: 0.8106 - val_loss: 0.6015 - val_accuracy: 0.7362
Epoch 4/10
200/200 [==============================] - 166s 833ms/step - loss: 0.4776 - accuracy: 0.7445 - val_loss: 0.6535 - val_accuracy: 0.7250
Epoch 5/10
200/200 [==============================] - 16s 80ms/step - loss: 0.3637 - accuracy: 0.8459 - val_loss: 0.5819 - val_accuracy: 0.7437
Epoch 6/10
200/200 [==============================] - 34s 168ms/step - loss: 0.3049 - accuracy: 0.8772 - val_loss: 0.6557 - val_accuracy: 0.7369
Epoch 7/10
```

200/200 [==============================] - 26s 132ms/step - loss: 0.2617 - accuracy: 0.9008 - val_loss: 0.7037 - val_accuracy: 0.7337
Epoch 8/10
200/200 [==============================] - 38s 189ms/step - loss: 0.2293 - accuracy: 0.9133 - val_loss: 0.7746 - val_accuracy: 0.7300
Epoch 9/10
200/200 [==============================] - 30s 152ms/step - loss: 0.1889 - accuracy: 0.9333 - val_loss: 0.8197 - val_accuracy: 0.7319
Epoch 10/10
200/200 [==============================] - 33s 168ms/step - loss: 0.1773 - accuracy: 0.9384 - val_loss: 0.8782 - val_accuracy: 0.7356

Meeting 9 Apr 23, 2024 5:30 PM

Domain1:2500 ,domain2:4000

```
Training Logistic Regression...
Logistic Regression Performance:
              precision    recall  f1-score   support

           0       0.81      0.83      0.82      1305
           1       0.78      0.76      0.77      1015

    accuracy                           0.80      2320
   macro avg       0.80      0.79      0.79      2320
weighted avg       0.80      0.80      0.80      2320

Training Support Vector Machine...
Support Vector Machine Performance:
              precision    recall  f1-score   support

           0       0.86      0.85      0.85      1305
           1       0.81      0.82      0.82      1015

    accuracy                           0.84      2320
   macro avg       0.83      0.84      0.84      2320
weighted avg       0.84      0.84      0.84      2320

Training Random Forest...
Random Forest Performance:
              precision    recall  f1-score   support

           0       0.84      0.93      0.88      1305
           1       0.89      0.78      0.83      1015
```

```
    accuracy                           0.86      2320
   macro avg       0.87      0.85      0.86      2320
weighted avg       0.87      0.86      0.86      2320

Training Neural Network...
Neural Network Performance:
              precision    recall  f1-score   support

           0       0.87      0.89      0.88      1305
           1       0.85      0.83      0.84      1015

    accuracy                           0.86      2320
   macro avg       0.86      0.86      0.86      2320
weighted avg       0.86      0.86      0.86      2320
```

Domain1:4000 ,domain2:4000

```
Training Logistic Regression...
Logistic Regression Performance:
              precision    recall  f1-score   support

           0       0.85      0.79      0.82      1319
           1       0.79      0.85      0.82      1241

    accuracy                           0.82      2560
   macro avg       0.82      0.82      0.82      2560
weighted avg       0.82      0.82      0.82      2560

Training Support Vector Machine...
Support Vector Machine Performance:
              precision    recall  f1-score   support

           0       0.87      0.84      0.86      1319
           1       0.83      0.87      0.85      1241

    accuracy                           0.85      2560
   macro avg       0.85      0.85      0.85      2560
weighted avg       0.86      0.85      0.85      2560

Training Random Forest...
Random Forest Performance:
              precision    recall  f1-score   support

           0       0.89      0.89      0.89      1319
           1       0.89      0.89      0.89      1241

    accuracy                           0.89      2560
   macro avg       0.89      0.89      0.89      2560
```

```
weighted avg       0.89      0.89      0.89      2560

Training Neural Network...
Neural Network Performance:
              precision    recall  f1-score   support

           0       0.89      0.88      0.89      1319
           1       0.88      0.89      0.88      1241

    accuracy                           0.88      2560
   macro avg       0.88      0.88      0.88      2560
weighted avg       0.88      0.88      0.88      2560
```

Domain1:3000 ,domain2:3000

```
Training Logistic Regression...
Logistic Regression Performance:
              precision    recall  f1-score   support

           0       0.82      0.77      0.79       992
           1       0.77      0.82      0.79       928

    accuracy                           0.79      1920
   macro avg       0.80      0.80      0.79      1920
weighted avg       0.80      0.79      0.79      1920

Training Support Vector Machine...
Support Vector Machine Performance:
              precision    recall  f1-score   support

           0       0.83      0.81      0.82       992
           1       0.80      0.83      0.81       928

    accuracy                           0.82      1920
   macro avg       0.82      0.82      0.82      1920
weighted avg       0.82      0.82      0.82      1920

Training Random Forest...
Random Forest Performance:
              precision    recall  f1-score   support

           0       0.84      0.86      0.85       992
           1       0.84      0.83      0.83       928

    accuracy                           0.84      1920
   macro avg       0.84      0.84      0.84      1920
weighted avg       0.84      0.84      0.84      1920
```

Training Neural Network...
Neural Network Performance:

	precision	recall	f1-score	support
0	0.84	0.87	0.85	992
1	0.85	0.82	0.84	928
accuracy			0.85	1920
macro avg	0.85	0.85	0.85	1920
weighted avg	0.85	0.85	0.85	1920

Classification Report for Random Forest:

	precision	recall	f1-score	support
0	0.80	0.77	0.78	778
1	0.79	0.81	0.80	822
accuracy			0.79	1600
macro avg	0.79	0.79	0.79	1600
weighted avg	0.79	0.79	0.79	1600

Best LSA-SVM Performance:

	precision	recall	f1-score	support
0	0.86	0.65	0.74	653
1	0.71	0.89	0.79	627
accuracy			0.77	1280
macro avg	0.78	0.77	0.76	1280
weighted avg	0.78	0.77	0.76	1280

SVM Performance:

	precision	recall	f1-score	support
0	0.81	0.74	0.77	653
1	0.75	0.82	0.78	627
accuracy			0.78	1280
macro avg	0.78	0.78	0.78	1280
weighted avg	0.78	0.78	0.78	1280

Classification Report for AdaBoost:

	precision	recall	f1-score	support

0	0.68	0.66	0.67	778
1	0.69	0.71	0.70	822
accuracy			0.68	1600
macro avg	0.68	0.68	0.68	1600
weighted avg	0.68	0.68	0.68	1600

Meeting 10 Apr 24, 2024 5:30 PM

- Data preprocessing Jijiang WEN rewrote to shrink down the size
- Write the model selection Jijiang WEN
- Retrain SVM models and merge to the data preprocessing paragraph Chang Xu
    - Use model trained on domain 1 to predict domain 2
- Insert the figures Chang Xu DONE
- Random Forest & Adaboost & Adaboost and SVM under ensemble methods Chang Xu
- Table of metrics and describe Deep Learning Models & put it in the overleaf Daniel Manzano Aguayo
-

Meeting 11 Apr 25, 2024 12:00 PM

Report

Report Link

https://www.overleaf.com/4447686724wbnbyzvbghkr#b1cabe

Report Structure (1200 words in total)

1. A brief description of the problem and introduction of any notation that you adopt in the report;

2. Description of your final approach(s) to the generation detection problem, the motivation and reasoning behind it, and why you think it performed well/not well in the competition; and

3. Any other **alternatives you considered and why you chose your final approach over these** (this may be in the form of empirical evaluation, but it must be to support your reasoning—examples like “method A, got accuracy 0.6 and method B, got accuracy 0.7, hence I use method B”, with no further explanation, will be marked down). Figures and/or tables should be used to better illustrate your results.

    a. Metadata model with both domain combined
    b. Metadata model trained on each domain separately
    c. RF
    d. Adaboost

-
- Data processing (How we deal with different topic of domain)
    - Data processBalance dataset
    - Tf-idf
    - Standscaler
-
- Model we used
    - Svm baseline
    - Meta model (improve)
    - Rf model (adaboost)
    - Deep learning (
    -
- Metrics we need
    - F1
    - Accuracy
    - Precision
    - Recall

4. A discussion on addressing the differences in performance of different methods in the two different domains. Provide your analysis and insights on how the domain may have affected your results, and discuss any strategies or techniques you employed to mitigate the impact of the two domains in your approach.
    a. SVM for each domain separately

```
Classification Report for Domain 1:
              precision    recall  f1-score   support

           0       0.82      0.73      0.78       500
           1       0.76      0.84      0.80       500

    accuracy                           0.79      1000
   macro avg       0.79      0.79      0.79      1000
weighted avg       0.79      0.79      0.79      1000

Classification Report for Domain 2:
              precision    recall  f1-score   support

           0       0.88      1.00      0.94      2289
           1       1.00      0.01      0.02       311

    accuracy                           0.88      2600
   macro avg       0.94      0.50      0.48      2600
weighted avg       0.90      0.88      0.83      2600

Evaluation of Domain 1 Model on Domain 2 Data:
              precision    recall  f1-score   support

           0       0.95      0.35      0.51      2289
           1       0.15      0.86      0.26       311

    accuracy                           0.41      2600
   macro avg       0.55      0.60      0.38      2600
weighted avg       0.85      0.41      0.48      2600

Evaluation of Domain 2 Model on Domain 1 Data:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67       500
           1       0.00      0.00      0.00       500

    accuracy                           0.50      1000
```

macro avg	0.25	0.50	0.33	1000
weighted avg	0.25	0.50	0.33	1000

While domain 1 data set has a balanced distribution, domain 2 has a highly imbalanced class distribution with a much higher proportion of machine generated data. Training two separate baseline SVM models on each domain of data resulting in the following two confusion matrices. Such an imbalanced nature of the dataset could potentially result in models with poorer generalization ability and overfitting on the majority class.

To mitigate these issues, data resampling and oversampling the minority class were used during the preprocessing stage. Different models were also used for each dataset.

Report

Introduction Chang Xu

Methodology / preprocessing / a bit about the 2 domains Jijiang WEN

Discussion

- Svm_baseline
-

- Differences in performance of different methods in the two different domains Chang Xu
- All approaches compare and contrast (**DOT POINTS ONLY FOR NOW**)
    - Rf model

Meta model A - Weak Models = 2 Support vector machines , trained domain 1 and domain 2
Meta Model = Multilayer Perceptron

	precision	recall	f1-score	support
0	0.78	0.73	0.75	495
1	0.75	0.79	0.77	505
accuracy			0.76	1000
macro avg	0.76	0.76	0.76	1000
weighted avg	0.76	0.76	0.76	1000

Meta model B - **Weak Models** = KNN, NaiveBayes, LogisticRegression, RandomForest , trained concatenated domains. **Meta Model** = Multilayer Perceptron

	precision	recall	f1-score	support
0	0.85	0.90	0.88	1018
1	0.89	0.83	0.86	982
accuracy			0.87	2000
macro avg	0.87	0.87	0.87	2000
weighted avg	0.87	0.87	0.87	2000

	precision	recall	f1-score	support
0	0.87	0.84	0.86	1319
1	0.83	0.87	0.85	1241

Logistic Regression Performance:

	precision	recall	f1-score	support
0	0.81	0.83	0.82	1305
1	0.78	0.76	0.77	1015
accuracy			0.80	2320
macro avg	0.80	0.79	0.79	2320
weighted avg	0.80	0.80	0.80	2320

Metrics/ Model	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score	Training Accuracy	Kaggle Accuracy
BaseLine SVC	85%	85%	85%	85%	76%
MLP					
LSTM	83%	83%	83%	83%	77%
Logistic Regression	80%	79%	79%	80%	73%
Random Forest	89%	89%	89%	89%	80%
Meta Model 1	76%	76%	76%	76%	67%
Meta Model 2	87%	87%	87%	87%	79.6%

Conclusion

- Our final choice of models

Reference

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9720154

3000 2500
domain1 0:2500 1:2500 5000
domain2 0:2500 1:2500 5000
Rf 81 ->83

Using all training data
1500 method generate more la**bel1 instances**
submission tfidf one 5 times
3-5 times for TOM

83 84

81

RF23_new2.csv Complete · Jijiang WEN · 2h ago	0.76450
RF23_new1.csv Complete · Jijiang WEN · 2h ago	0.75700
MLPPytorchs8.csv Complete · ID1095549 · 4h ago	0.74250
MLP23_1.csv Complete · Jijiang WEN · 6h ago	0.75050

Draft 01

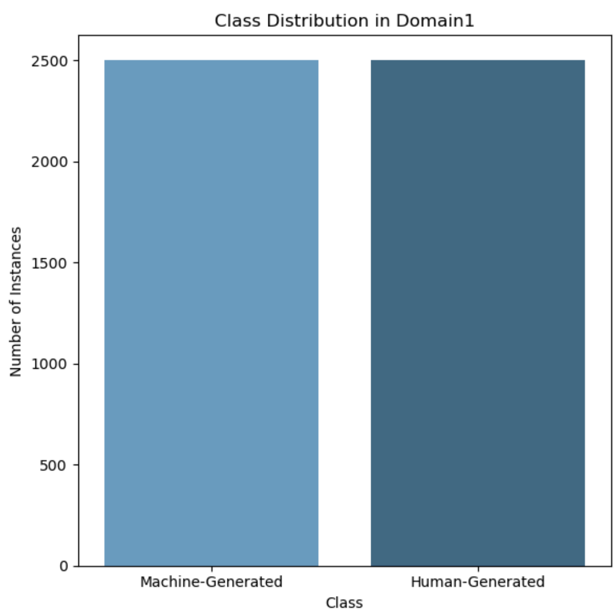
1.simple introduction

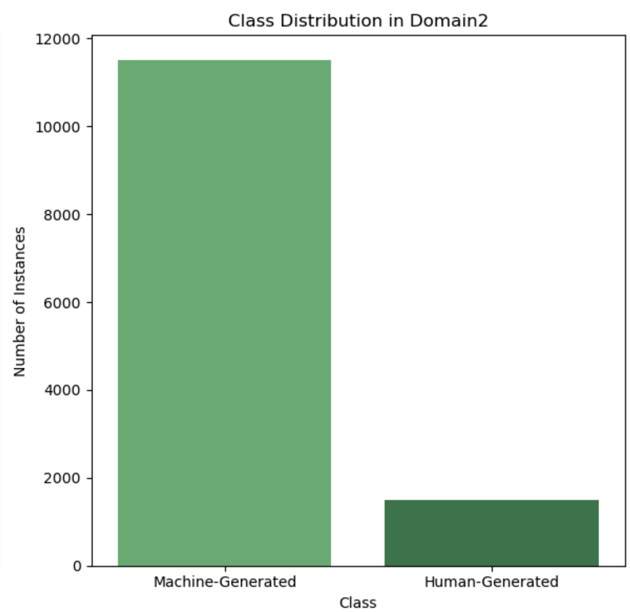
With the advancement of technology, text generation using over-parameterized models has a wide range of applications across various industries, such as chat bots, and content creation. However, this also poses a challenge in developing text generation detection models to ensure the authenticity and accuracy of areas like news reporting and to prevent the spread of fake news.

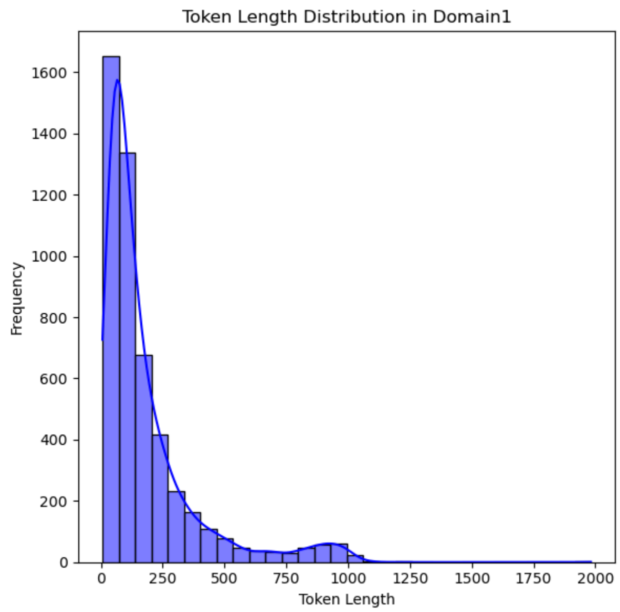
The following are a list of notations adopted in the report: ….

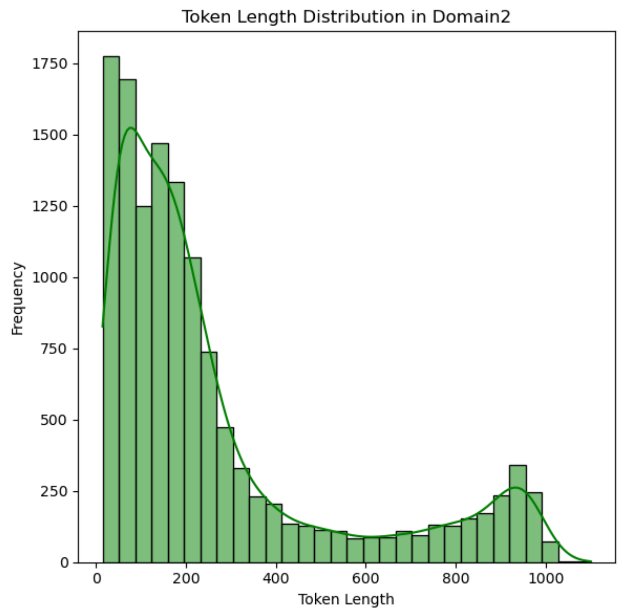
2.data process

We have two training datasets from distinct domains, each containing a mix of human-generated and machine-generated text data. label 0 represents machine-generated text data, and label 1 represents human-generated text data. Domain1 dataset is already balanced, In "domain 1", there are 2500 instances of both label 0 and label 1. In "domain2", there are 1500 instances of label 1 and 11,500 instances of label 0.









So we balanced domain2 dataset first,we oversampling human-generated instances to 2600, and undersampling machine-generated instances to 2600, Similarly, we oversampled each label in domain1 to 2600 instances. Additionally, to enable the model to handle datasets from different domains, we noted that some instances in domain1 exceeded 1,000 in length, whereas in domain2, instances rarely did. Therefore, we first removed instances from domain1 that were over 1,000 in length to make their features appear similar, and then merged them for training. We also used TF-IDF vectorization to highlight unique words in texts, improving our model's ability to distinguish between human and machine-generated content.

Classification Report for SVM Model - Dataset 1:

	precision	recall	f1-score	support
0	0.82	0.73	0.78	500
1	0.76	0.84	0.80	500
accuracy			0.79	1000
macro avg	0.79	0.79	0.79	1000
weighted avg	0.79	0.79	0.79	1000

Classification Report for SVM Model - Dataset 2:

	precision	recall	f1-score	support
0	0.88	1.00	0.94	2289
1	1.00	0.00	0.01	311

```
    accuracy                           0.88      2600
   macro avg       0.94      0.50      0.47      2600
weighted avg       0.89      0.88      0.83      2600
```

While domain 1 data set has a balanced distribution, domain 2 has a highly imbalanced class distribution with a much higher proportion of machine generated data. Training two separate baseline SVM models on each domain of data resulting in the following two confusion matrices. Such an imbalanced nature of the dataset could potentially result in models with poorer generalization ability and overfitting on the majority class.

To mitigate these issues, data resampling and oversampling the minority class were used during the preprocessing stage. Different models were also used for each dataset.

3.Baseline model

For the baseline model we use SVM, SVM is a classic binary classification machine learning model that is well suited for tasks such as distinguishing human-generated text and machine-generated text. The choice of RBF kernel is particularly suitable for this task as it allows the SVM to capture complex non-linear relationships between features extracted from text data. Using it as a baseline model also helps with comparing other models in subsequent analyses. The performance of SVM model for combined domain1 and domain2 processed dataset.

```
SVM Performance:
              precision    recall  f1-score   support

           0       0.82      0.78      0.80       788
           1       0.79      0.83      0.81       812

    accuracy                           0.81      1600
   macro avg       0.81      0.81      0.81      1600
weighted avg       0.81      0.81      0.81      1600
```

4.

Metadata model with both domain combined

The first approach is an ensemble learning technique. Stacking, also known as stacked generalization or stacking ensemble, is a powerful technique in machine learning that combines multiple base models and a meta model to improve the predictive performance.

For this task, the datasets from both domains are concatenated together (domain 1 and 2) in one dataset. The different text codes/characters are vectorized by applying the model TF-IDF as a feature extractor that works as input for the models. Additionally, the combined training

dataset (domain 1 and 2) is splitted in training and validation sets for the model's hyperparameter tuning.
The base/weak models employed in the Stack are the Logistic Regression, K-Nearest Neighbors, Random Forest, Naive Bayes, and the meta model is a Multilayer Perceptron (MLP).
The main focus on the tuning is to find the best hyperparameters for the Multi Layer Perceptron, the meta model. Therefore, heuristically, we attempt different activation functions such as Identity, Logistic, Tanh and ReLu. In the same way, different optimization algorithms are used such as lbfgs, sgd, adam.
The model accuracy obtained is 79.5% accuracy on the validation set, and 79.6% in the test set (uploaded in Kaggle).
However, the model performance is not better than a single model approach such as a RandomForest model. Therefore, in terms of efficiency this ensemble learning approach is more computationally expensive.

Metadata model trained on each domain separately

The second approach is slightly different from the first approach. The data from domain 1 is down, but the base/weak models are two support vector machines and a Multilayer Perceptron as a meta model.

Research Part

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9720154

5.deep learning model

- Ensemble method, random forest was adopted to detect the text generation.

- The test results on kaggle and the result obtained from the validation set differed by 10%, thus text augmentation that generates random noise in the training set was applied to mitigate the impact of overfitting.
- Another ensemble method, adaptive boosting, that focuses on instances that were misclassified by previous weak learners, was also adopted.

Both Random Forest that focus on combining multiple decision trees to make predictions and Adaptive Boosting that focus on improving the performance of weaker learners by sequentially emphasizing misclassified instances were adopted. The initial test results on the Kaggle and the result obtained from the validation set were quite different, thus the data augmentation was added to introduce random noise to help improve the generalization ability of the models. From the table [], though both models had their hyperparameters tuned via grid search, the Random Forest model had a better performance across all the metrics. This could be attributed to the Adaptive Boosting focuses more on the nuances of the training set which could result in its poor generalization ability to the unseen data.

```
Epoch 1
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.63      0.28      0.39       983
         1.0       0.55      0.84      0.66      1017

    accuracy                           0.57      2000
   macro avg       0.59      0.56      0.53      2000
weighted avg       0.59      0.57      0.53      2000

Test Error:
 Accuracy: 56.5%, Avg loss: 0.680076

Epoch 2
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.68      0.66      0.67       983
         1.0       0.68      0.70      0.69      1017

    accuracy                           0.68      2000
   macro avg       0.68      0.68      0.68      2000
```

```
weighted avg       0.68      0.68      0.68      2000

Test Error:
 Accuracy: 68.0%, Avg loss: 0.590462

Epoch 3
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.79      0.66      0.72       983
         1.0       0.72      0.83      0.77      1017

    accuracy                           0.75      2000
   macro avg       0.75      0.75      0.74      2000
weighted avg       0.75      0.75      0.74      2000

Test Error:
 Accuracy: 74.7%, Avg loss: 0.533163

Epoch 4
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.75      0.85      0.79       983
         1.0       0.83      0.72      0.77      1017

    accuracy                           0.78      2000
   macro avg       0.79      0.78      0.78      2000
weighted avg       0.79      0.78      0.78      2000

Test Error:
 Accuracy: 78.3%, Avg loss: 0.534718

Epoch 5
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.80      0.83      0.81       983
         1.0       0.83      0.79      0.81      1017

    accuracy                           0.81      2000
   macro avg       0.81      0.81      0.81      2000
weighted avg       0.81      0.81      0.81      2000
```

```
Test Error: 
 Accuracy: 81.3%, Avg loss: 0.501408 

Epoch 6
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.78      0.87      0.82       983
         1.0       0.86      0.77      0.81      1017

    accuracy                           0.82      2000
   macro avg       0.82      0.82      0.82      2000
weighted avg       0.82      0.82      0.82      2000

Test Error: 
 Accuracy: 81.7%, Avg loss: 0.504244 

Epoch 7
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.79      0.88      0.84       983
         1.0       0.87      0.78      0.82      1017

    accuracy                           0.83      2000
   macro avg       0.83      0.83      0.83      2000
weighted avg       0.83      0.83      0.83      2000

Test Error: 
 Accuracy: 82.9%, Avg loss: 0.547115 

Epoch 8
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.88      0.84       983
         1.0       0.87      0.80      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error: 
 Accuracy: 84.0%, Avg loss: 0.533395 
```

```
Epoch 9
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.82      0.87      0.85       983
         1.0       0.87      0.82      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.85      0.84      0.84      2000
weighted avg       0.85      0.84      0.84      2000

Test Error:
 Accuracy: 84.5%, Avg loss: 0.563532

Epoch 10
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.85      0.85      0.85       983
         1.0       0.86      0.85      0.85      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 85.2%, Avg loss: 0.573540

Epoch 11
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.88      0.84       983
         1.0       0.87      0.80      0.83      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error:
 Accuracy: 83.5%, Avg loss: 0.557771

Epoch 12
```

```
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.80      0.87      0.83       983
         1.0       0.86      0.79      0.83      1017

    accuracy                           0.83      2000
   macro avg       0.83      0.83      0.83      2000
weighted avg       0.83      0.83      0.83      2000

Test Error: 
 Accuracy: 83.0%, Avg loss: 0.596245 

Epoch 13
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.85      0.86      0.85       983
         1.0       0.86      0.85      0.86      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error: 
 Accuracy: 85.4%, Avg loss: 0.541854 

Epoch 14
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      0.86      0.84       983
         1.0       0.86      0.82      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error: 
 Accuracy: 84.2%, Avg loss: 0.562143 

Epoch 15
-------------------------------
Classification Report:
```

```
              precision    recall  f1-score   support

         0.0       0.82      0.87      0.85       983
         1.0       0.87      0.82      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.85      0.84      0.84      2000
weighted avg       0.85      0.84      0.84      2000

Test Error:
 Accuracy: 84.5%, Avg loss: 0.613867

Epoch 16
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.82      0.87      0.85       983
         1.0       0.87      0.82      0.84      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 84.5%, Avg loss: 0.576971

Epoch 17
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.88      0.84       983
         1.0       0.87      0.80      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error:
 Accuracy: 84.0%, Avg loss: 0.590549

Epoch 18
-------------------------------
Classification Report:
              precision    recall  f1-score   support
```

```
             0.0       0.80      0.88      0.84       983
             1.0       0.88      0.79      0.83      1017

        accuracy                           0.84      2000
       macro avg       0.84      0.84      0.84      2000
    weighted avg       0.84      0.84      0.84      2000

Test Error:
 Accuracy: 83.6%, Avg loss: 0.567819

Epoch 19
-------------------------------
Classification Report:
                  precision    recall  f1-score   support

             0.0       0.81      0.87      0.84       983
             1.0       0.87      0.81      0.84      1017

        accuracy                           0.84      2000
       macro avg       0.84      0.84      0.84      2000
    weighted avg       0.84      0.84      0.84      2000

Test Error:
 Accuracy: 83.9%, Avg loss: 0.601454

Epoch 20
-------------------------------
Classification Report:
                  precision    recall  f1-score   support

             0.0       0.83      0.86      0.85       983
             1.0       0.86      0.83      0.85      1017

        accuracy                           0.85      2000
       macro avg       0.85      0.85      0.85      2000
    weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 84.8%, Avg loss: 0.562080

Epoch 21
-------------------------------
Classification Report:
                  precision    recall  f1-score   support

             0.0       0.82      0.87      0.84       983
             1.0       0.87      0.81      0.84      1017
```

```
    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error: 
 Accuracy: 84.0%, Avg loss: 0.575698 

Epoch 22
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.88      0.84       983
         1.0       0.87      0.81      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error: 
 Accuracy: 84.2%, Avg loss: 0.574204 

Epoch 23
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.84      0.87      0.86       983
         1.0       0.87      0.84      0.86      1017

    accuracy                           0.86      2000
   macro avg       0.86      0.86      0.86      2000
weighted avg       0.86      0.86      0.86      2000

Test Error: 
 Accuracy: 85.5%, Avg loss: 0.603384 

Epoch 24
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.82      0.89      0.85       983
         1.0       0.88      0.82      0.85      1017

    accuracy                           0.85      2000
```

```
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 85.1%, Avg loss: 0.579978

Epoch 25
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      0.88      0.85       983
         1.0       0.88      0.83      0.85      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 85.3%, Avg loss: 0.602619

Epoch 26
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.82      0.87      0.85       983
         1.0       0.87      0.82      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.85      0.84      0.84      2000

Test Error:
 Accuracy: 84.4%, Avg loss: 0.603330

Epoch 27
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.90      0.85       983
         1.0       0.89      0.79      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.85      0.85      0.84      2000
weighted avg       0.85      0.84      0.84      2000
```

```
Test Error:
 Accuracy: 84.5%, Avg loss: 0.584491

Epoch 28
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.82      0.88      0.85       983
         1.0       0.88      0.81      0.84      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 84.8%, Avg loss: 0.622151

Epoch 29
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      0.88      0.85       983
         1.0       0.88      0.82      0.85      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 85.0%, Avg loss: 0.596417

Epoch 30
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.80      0.89      0.85       983
         1.0       0.88      0.79      0.83      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error:
```

```
 Accuracy: 84.0%, Avg loss: 0.651260

Epoch 31
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.88      0.84       983
         1.0       0.88      0.80      0.84      1017

    accuracy                           0.84      2000
   macro avg       0.84      0.84      0.84      2000
weighted avg       0.84      0.84      0.84      2000

Test Error:
 Accuracy: 84.0%, Avg loss: 0.649983

Epoch 32
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      0.89      0.86       983
         1.0       0.88      0.82      0.85      1017

    accuracy                           0.85      2000
   macro avg       0.85      0.85      0.85      2000
weighted avg       0.85      0.85      0.85      2000

Test Error:
 Accuracy: 85.3%, Avg loss: 0.613521

Epoch 33
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.86      0.88      0.87       983
         1.0       0.88      0.86      0.87      1017

    accuracy                           0.87      2000
   macro avg       0.87      0.87      0.87      2000
weighted avg       0.87      0.87      0.87      2000

Test Error:
 Accuracy: 87.0%, Avg loss: 0.570445
```

```
Epoch 34
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.85      0.88      0.87       983
         1.0       0.88      0.85      0.87      1017

    accuracy                           0.87      2000
   macro avg       0.87      0.87      0.87      2000
weighted avg       0.87      0.87      0.87      2000

Test Error:
 Accuracy: 86.6%, Avg loss: 0.598824

Epoch 35
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.85      0.87      0.86       983
         1.0       0.87      0.85      0.86      1017

    accuracy                           0.86      2000
   macro avg       0.86      0.86      0.86      2000
weighted avg       0.86      0.86      0.86      2000

Test Error:
 Accuracy: 86.0%, Avg loss: 0.595112

Done!
final test:
```

```
Epoch 1
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.91      0.88      0.89       983
         1.0       0.89      0.91      0.90      1017

    accuracy                           0.90      2000
   macro avg       0.90      0.90      0.90      2000
weighted avg       0.90      0.90      0.90      2000
```

```
Test Error: 
 Accuracy: 89.8%, Avg loss: 0.277091 

Epoch 2
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.91      0.91      0.91       983
         1.0       0.92      0.91      0.91      1017

    accuracy                           0.91      2000
   macro avg       0.91      0.91      0.91      2000
weighted avg       0.91      0.91      0.91      2000

Test Error: 
 Accuracy: 91.1%, Avg loss: 0.213115 

Epoch 3
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      0.89      0.91       983
         1.0       0.90      0.94      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.91      0.91      2000
weighted avg       0.92      0.92      0.91      2000

Test Error: 
 Accuracy: 91.5%, Avg loss: 0.214450 

Epoch 4
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      0.90      0.91       983
         1.0       0.91      0.93      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.92      0.92      2000
weighted avg       0.92      0.92      0.92      2000

Test Error: 
 Accuracy: 91.6%, Avg loss: 0.224305 
```

```
Epoch 5
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      0.90      0.91       983
         1.0       0.91      0.93      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.92      0.92      2000
weighted avg       0.92      0.92      0.92      2000

Test Error:
 Accuracy: 91.8%, Avg loss: 0.234717

Epoch 6
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.92      0.90      0.91       983
         1.0       0.91      0.93      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.92      0.92      2000
weighted avg       0.92      0.92      0.92      2000

Test Error:
 Accuracy: 91.6%, Avg loss: 0.243511

Epoch 7
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.91      0.91      0.91       983
         1.0       0.92      0.91      0.91      1017

    accuracy                           0.91      2000
   macro avg       0.91      0.91      0.91      2000
weighted avg       0.91      0.91      0.91      2000

Test Error:
 Accuracy: 91.3%, Avg loss: 0.254917

Epoch 8
```

```
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      0.89      0.91       983
         1.0       0.90      0.94      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.91      0.91      2000
weighted avg       0.92      0.92      0.91      2000

Test Error:
 Accuracy: 91.5%, Avg loss: 0.266437

Epoch 9
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      0.90      0.91       983
         1.0       0.90      0.93      0.92      1017

    accuracy                           0.92      2000
   macro avg       0.92      0.92      0.92      2000
weighted avg       0.92      0.92      0.92      2000

Test Error:
 Accuracy: 91.5%, Avg loss: 0.273246

Epoch 10
-------------------------------
Classification Report:
              precision    recall  f1-score   support

         0.0       0.90      0.92      0.91       983
         1.0       0.92      0.90      0.91      1017

    accuracy                           0.91      2000
   macro avg       0.91      0.91      0.91      2000
weighted avg       0.91      0.91      0.91      2000

Test Error:
 Accuracy: 91.0%, Avg loss: 0.285266

Done!
final test:
```