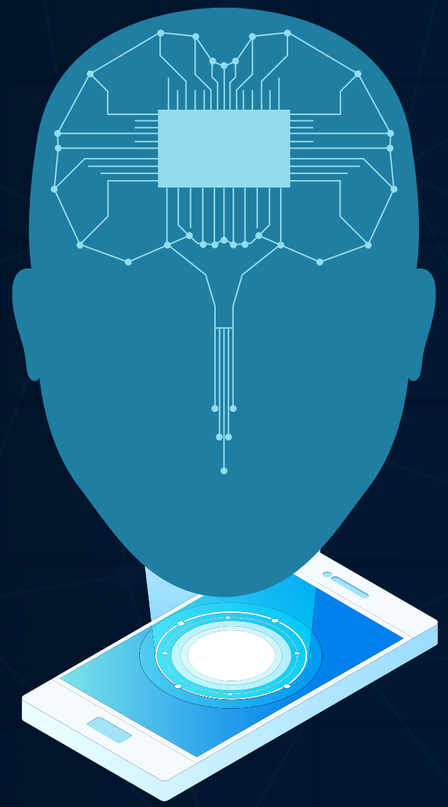


INTENT DETECTION THROUGH A NEURAL NETWORK-ENABLED TEXT CLASSIFIER

GROUP 2

- ALEXANDER HEGER
- MUGDHA KHAIRNAR
- ISHAN NAGRANI

AGENDA

01 Problem Statement

02 Data Preparation for NLP

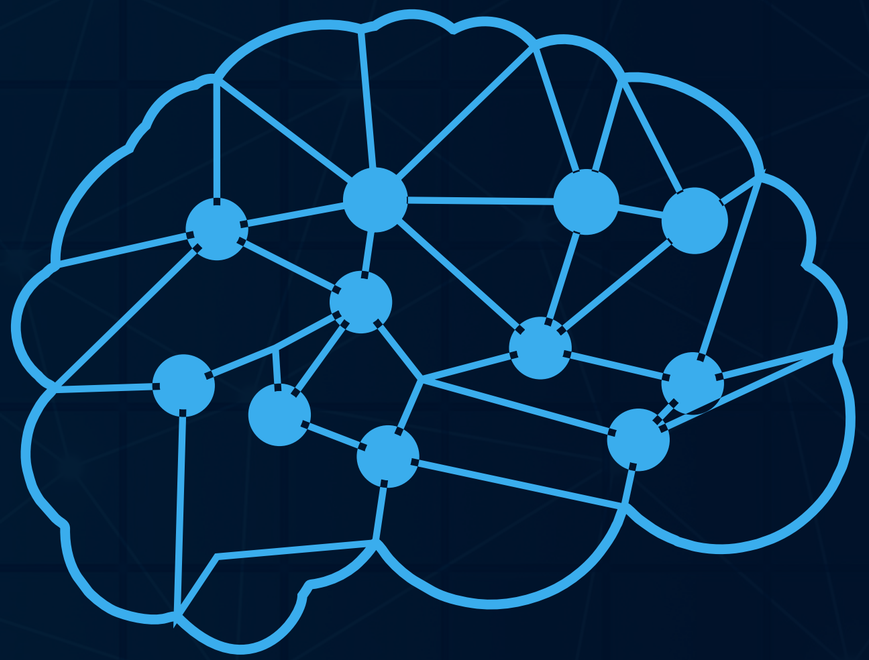
03 Model Development

04 Conclusions

1

PROBLEM STATEMENT

Our objective was to develop a neural network text classifier that accurately predicted the intent of Stack Overflow posts, and to identify the best model for prediction.



2

MOTIVATION

We selected this task to imitate the initial stages of a chatbot powered by a language model, an applicable tool to many industries with an online presence.

OUR DATASET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 139638 entries, 0 to 139637
Data columns (total 20 columns):
 #   Column                   Non-Null Count   Dtype
---  ------                   --------------   -----
 0   id                       139638 non-null  int64
 1   title                    139638 non-null  object
 2   body                     139638 non-null  object
 3   accepted_answer_id       41278 non-null   float64
 4   answer_count             139638 non-null  int64
 5   comment_count            139638 non-null  int64
 6   community_owned_date     0 non-null       float64
 7   creation_date            139638 non-null  object
 8   favorite_count           8157 non-null    float64
 9   last_activity_date       139638 non-null  object
 10  last_edit_date           64703 non-null   object
 11  last_editor_display_name 231 non-null     object
 12  last_editor_user_id      64472 non-null   float64
 13  owner_display_name       612 non-null     object
 14  owner_user_id            139053 non-null  float64
 15  parent_id                0 non-null       float64
 16  post_type_id             139638 non-null  int64
 17  score                    139638 non-null  int64
 18  tags                     139638 non-null  object
 19  view_count               139638 non-null  int64
dtypes: float64(6), int64(6), object(8)
```

SOURCE

Stack Exchange Data Explorer
Hosted on GCP BigQuery

DATA PREPROCESSING

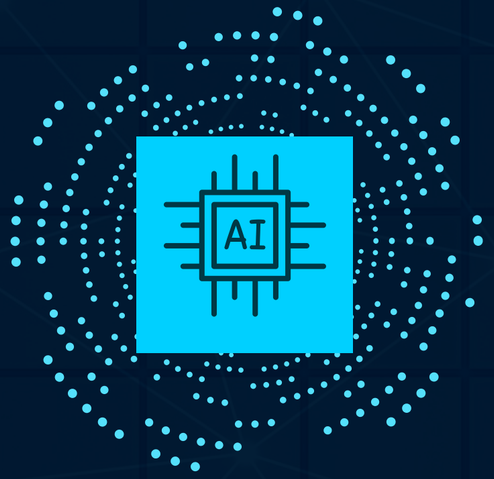
1. Merge title and body in order to make a combined text feature, our corpus.

2. Remove symbols and extra spaces from text.

3. Remove stopwords found in spaCy's English stopword dictionary.

DENSE NETWORK

python	7854
javascript	5417
java	2293
c#	2153
reactjs	1728
r	1422
flutter	1356
android	1313
php	1130
c++	1113



PREPROCESSING

- 25,779 documents with top 10 labels
- 10,000 most frequent tokens
- Unbalanced class weights

TOPOLOGY

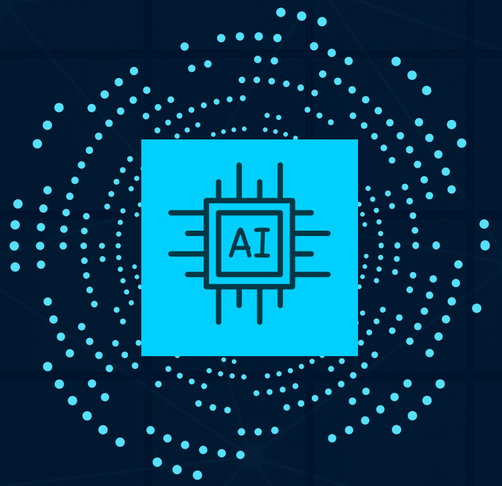
- One-hot encoding
- Dense with L2 regularization
- Dropout

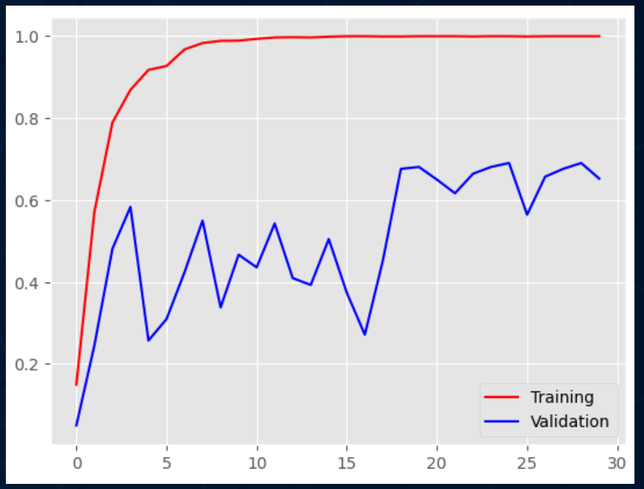
PERFORMANCE

- Categorical Accuracy:
  - Best val: 0.8419
  - Test set: 0.8357

CONVOLUTIONAL NEURAL NETWORK

python	822
javascript	532
java	234
c#	215
android	155
reactjs	153
flutter	143
r	136
c++	118
php	117





PREPROCESSING

- 2,625 documents with top 10 labels
- TextVectorization with 2,000 tokens
- Balanced class weights

TOPOLOGY

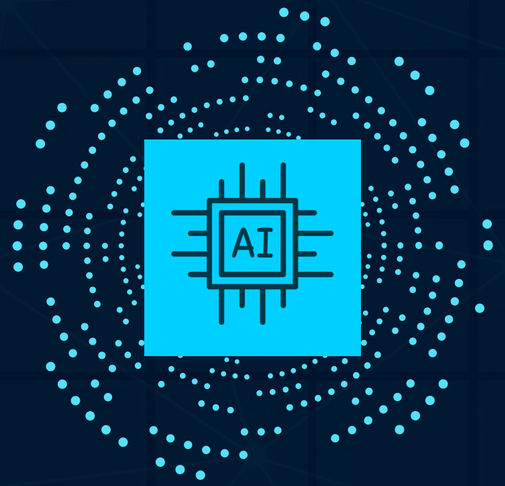
- One-hot encoding
- SeparableConv1D
- BatchNormalization
- GlobalAveragePooling

PERFORMANCE

- Categorical Accuracy:
  - Best val: 0.6905
  - Test set: 0.7048

RECURRENT NEURAL NETWORK

python	822
javascript	532
java	234
c#	215
android	155
reactjs	153
flutter	143
r	136
c++	118
php	117



PREPROCESSING

- Same 2,625 documents
- TextVectorization with 2,000 tokens
- Balanced class weights

TOPOLOGY

- 8-dimensional word embeddings
- Bidirectional LSTM
- Dropout
- BatchNormalization

PERFORMANCE

- Categorical Accuracy:
  - Best val: 0.6643
  - Test set: 0.6267

BERT LANGUAGE MODEL

MODEL OVERVIEW

PREPROCESSING

- Removed html tags, extra spaces
- Converted labels to numeric and one-hot encoded

TOPOLOGY

- Input Layers
  - Tokenised text and attention masks; Max length of 256
- Pre-trained BERT layer
  - Pooled output of 768 dimension containing embeddings
- Intermediate layer
  - Dense layers with 512 units, relu activation
- Output layer
  - Dense layer with 10 units, softmax activation

TRANSFORMER

'bert-base-case' pre-trained model
12 layers, 768 hidden units and 110M Parameters

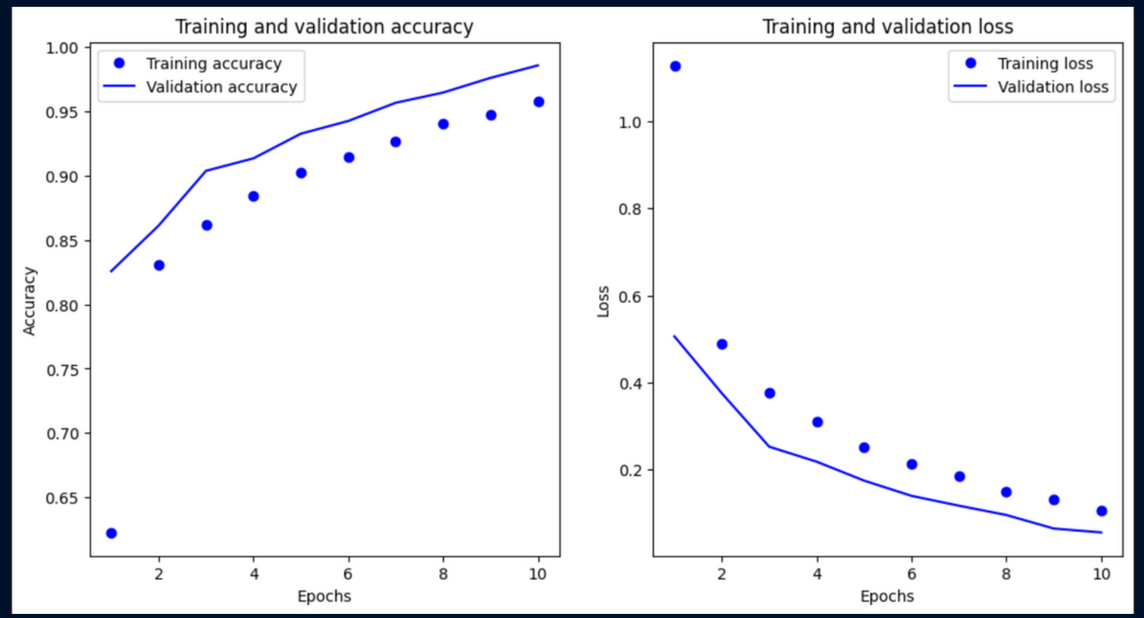
```
Model: "model"
__________________________________________________________________________________________________
 Layer (type)                   Output Shape              Param #     Connected to
==================================================================================================
 input_ids (InputLayer)         [(None, 256)]             0           []

 attention_mask (InputLayer)    [(None, 256)]             0           []

 bert (TFBertMainLayer)         TFBaseModelOutputWi       108310272   ['input_ids[0][0]',
                                thPoolingAndCrossAt                    'attention_mask[0][0]']
                                tentions(last_hidde
                                n_state=(None, 256,
                                 768),
                                 pooler_output=(Non
                                e, 768),
                                 past_key_values=No
                                ne, hidden_states=N
                                one, attentions=Non
                                e, cross_attentions
                                =None)

 intermediate_layer (Dense)     (None, 512)               393728      ['bert[0][1]']

 output_layer (Dense)           (None, 10)                5130        ['intermediate_layer[0][0]']

==================================================================================================
Total params: 108,709,130
Trainable params: 108,709,130
Non-trainable params: 0
```

BERT LANGUAGE MODEL

MODEL RESULTS

HYPERPERAMETERS

- Optimizer
  - Adam (learning_rate=1e-5)
- Loss function
  - Categorical Crossentropy
- Batch Size
  - 16
- Epochs
  - 10



PERFORMANCE

- Validation Accuracy
  - Achieved roughly 99% accuracy
- Validation Loss
  - Minimised Validation loss to 0.05

BERT LANGUAGE MODEL

PREDICTION EXAMPLES

I can't install any package using pip or pipenv

Asked today Modified today Viewed 32 times

-1

I want to install many packages for my project like flask or psycopg2 or flask-sqlachermy but i can't. When i use the pip i receive this error message:

```
**ERROR: Could not install packages due to an OSError: [Errno 13] Perm
Consider using the `--user` option or check the permissions.**
When i use the pipenv i receive this error message:
**[36mERROR: Could not install packages due to an OSError: [Errno 13]
'C:\\Users\\ASUS\\.virtualenvs\\ASUS-mVsQ382S\\Lib\\site-packages\\mar
Check the permissions.**
```

[enter image description here](https://i.stack.imgur.com/Ic1NB.png)

I tried to install again the markupsafe package because i thint that it is the source of the problem but it's not working

```
input_text = input('Enter coding question here: ')
processed_data = prepare_data(input_text, tokenizer)
result = make_prediction(sentiment_model, processed_data=processed_data)
print(f"Predicted Sentiment: {result}")
```

```
Enter coding question here: I can't install any package using pip or pipenv.
1/1 [==============================] - 0s 62ms/step
Predicted Sentiment: Python
```

How can I prevent SQL injection in PHP?

Asked 14 years, 7 months ago Modified 4 months ago Viewed 2.1m times

2773

This question's answers are a community effort. Edit existing answers to improve this post. It is not currently accepting new answers or interactions.

If user input is inserted without modification into an SQL query, then the application becomes vulnerable to SQL injection, like in the following example:

```
$unsafe_variable = $_POST['user_input'];
mysql_query("INSERT INTO `table` (`column`) VALUES ('$unsafe_variable
```

That's because the user can input something like `value'); DROP TABLE table;--`, and the query becomes:

```
INSERT INTO `table` (`column`) VALUES('value'); DROP TABLE table;--'
```

What can be done to prevent this from happening?

```
input_text = input('Enter coding question here: ')
processed_data = prepare_data(input_text, tokenizer)
result = make_prediction(sentiment_model, processed_data=processed_data)
print(f"Predicted Sentiment: {result}")
```

```
Enter coding question here: How can I prevent SQL injection in PHP? If user
1/1 [==============================] - 0s 57ms/step
Predicted Sentiment: php
```

CHALLENGES & FUTURE SCOPE

1. USE PRE-TRAINED TRANSFORMRS WHEN FEASIBLE
2. GARBAGE IN, GARBAGE OUT
3. FEATURE SELECTION
4. PREDICTING TOP N LABELS
5. CHATBOT IMPLEMENTATION

THANKS!

Do you have questions?