# Prediction of Stack Overflow Tag

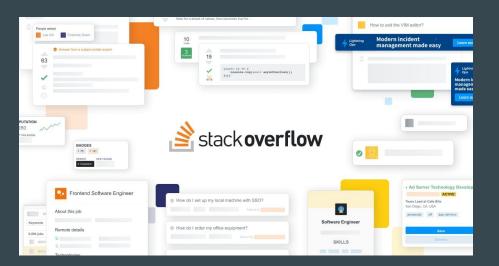
 $\bullet \bullet \bullet$ 

NLP Custom Project Prajwal Khot

#### What is Stack Overflow?

Stack Overflow is a question and answer website for professional and enthusiast programmers.

The website serves as a platform for users to ask and answer questions, and, through membership and active participation, to vote questions and answers up or down similar to Reddit and edit questions and answers in a fashion similar to a wiki.



#### **Motivation**

I won't be lying if I assert that every developer/engineer/student has used the website Stack Overflow more than once in their journey. Widely considered as one of the largest and more trusted websites for developers to learn and share their knowledge, the website presently hosts in excess of 10,000,000 questions.

A tag is a word or phrase that describes the topic of the question. Tags are a means of connecting experts with questions they will be able to answer by sorting questions into specific, well-defined categories.

#### **Spacy and Stopwords**

 Stop words are those words in natural language that have a very little meaning, such as "is", "an", "the", etc.

 Stop words are often removed from the text before training deep learning and machine learning models since stop words occur in abundance, hence providing little to no unique information that can be used for classification or clustering.



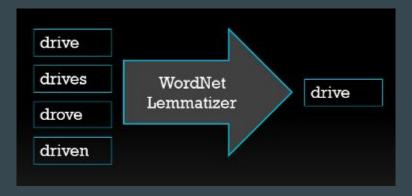
### Spacy and Stopwords

- In Python, there are myriad of options to use in order to remove stop words. Some of the libraries are NLTK, SpaCy, Gensim, TextBlob.
- In this project, I used spaCy library.



#### **Word Lemmatization**

- Lemmatization is the process of converting a word to its base form.
- The difference between stemming and lemmatization is, lemmatization considers
  the context and converts the word to its meaningful base form, whereas stemming
  just removes the last few characters, often leading to incorrect meanings and
  spelling errors.
- In this project I used Wordnet Lemmatizer by NLTK



# Sample Data

2008-08-01T23:22:08Z

#### Question:

4 260

| Ids | <b>OwnerUserId</b> | CreationDate         | ClosedDate           | Score | Title  | Body   |
|-----|--------------------|----------------------|----------------------|-------|--|--|
| 80  | 26.0               | 2008-08-01T13:57:07Z | NaN                  | 26    | SQLStatement.execute() - multiple queries in o | I've written a database generation script i          |
| 90  | 58.0               | 2008-08-01T14:41:24Z | 2012-12-26T03:45:49Z | 144   | Good branching and merging tutorials for Torto | Are there any really good tutorials explain          |
| 120 | 83.0               | 2008-08-01T15:50:08Z | NaN                  | 21    | ASP.NET Site Maps                              | Has anyone got experience creating <strong></strong> |
| 180 | 2089740.0          | 2008-08-01T18:42:19Z | NaN                  | 53    | Function for creating color wheels             | This is something I've pseudo-solved many t          |

Adding scripting functionality to .NET applica...

NaN

49

I have a little game written in C#. It uses...

# Sample Data

#### Answers:

|   | Id  | A_Score | A_Body   |
|---|-----|---------|--|
| 0 | 90  | 13      | <a href="http://svnbook.red-bean.com/">Vers</a>              |
| 1 | 80  | 12      | I wound up using this. It is a kind of a ha                  |
| 2 | 180 | 1       | I've read somewhere the human eye can't dis                  |
| 3 | 260 | 4       | Yes, I thought about that, but I soon figur                  |
| 4 | 260 | 28      | <a <="" href="http://www.codeproject.com/Article" th=""></a> |

#### Tags:

| Tag            | Id |   |
|----------------|----|---|
| flex           | 80 | 0 |
| actionscript-3 | 80 | 1 |
| air            | 80 | 2 |
| svn            | 90 | 3 |
| tortoisesvn    | 90 | 4 |

# Cleaning and Combining the dataset

[the, difference, between, a, datagrid, and, a...

|    | Title  | Body   | Tags           |
|----|--|--|----------------|
| 11 | [how, to, get, the, value, of, built, encoded, | I need to grab the base64-encoded representati | c# asp.net     |
| 19 | [can, i, logically, reorder, columns, in, a, t | If I'm adding a column to a table in Microsoft | sql-server     |
| 23 | [convert, hashbytes, to, varchar]              | I want to get the MD5 Hash of a string value i | sql sql-server |
| 34 | [mysqlapache, error, in, php, mysql, query]    | I am getting the following error:\n\n\n Acces  | php mysql      |
|    |  |  |                |

I've been doing ASP.NET development for a litt...

asp.net

#### **TF-IDF Vectorization**

- TF-IDF is an abbreviation for Term Frequency Inverse Document Frequency.
- It is used to convert a collection of raw documents to a matrix of TF-IDF features.
- It is the measure of originality of a word by comparing the number of times a word appears in a document with the number of documents it appears in.
- Here each question is considered as a document.

#### Categorical Encoding

- Few algorithms such as CATBOAST, decision-trees can handle categorical values very well but most of the algorithms expect numerical values to achieve state-of-the-art results.
- There are 2 main methods: One-Hot-Encoding and Label-Encoder
- In this project I used LabelEncoder method

### **Categorical Encoding**

#### Label Encoding

| Food Name | Categorical # | Calories |
|-----------|---------------|----------|
| Apple     | 1             | 95       |
| Chicken   | 2             | 231      |
| Broccoli  | 3             | 50       |

#### One Hot Encoding

| Apple | Chicken | Broccoli | Calories |
|-------|---------|----------|----------|
| 1     | 0       | 0        | 95       |
| 0     | 1       | 0        | 231      |
| 0     | 0       | 1        | 50       |

#### Classifiers Used

- Logistic Regression
- XGB Classifier
- Multinomial Naive Bayes
- K-Nearest Neighbors(KNN)
- Random Forest Classifier

# **Logistic Regression**

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.62      | 0.97   | 0.75     | 203     |
| 1            | 0.62      | 0.11   | 0.18     | 47      |
| 2            | 0.93      | 0.45   | 0.60     | 29      |
| 3            | 0.63      | 0.58   | 0.61     | 62      |
| 4            | 0.34      | 0.46   | 0.39     | 125     |
| 5            | 0.56      | 0.15   | 0.23     | 34      |
| 6            | 0.67      | 0.67   | 0.67     | 127     |
| 7            | 0.53      | 0.31   | 0.40     | 51      |
| 8            | 0.88      | 0.54   | 0.67     | 13      |
| 9            | 1.00      | 0.96   | 0.98     | 80      |
| 10           | 0.44      | 0.73   | 0.55     | 70      |
| 11           | 1.00      | 0.31   | 0.47     | 29      |
| 12           | 0.61      | 0.73   | 0.66     | 158     |
| 13           | 0.33      | 0.04   | 0.07     | 27      |
| 14           | 0.48      | 0.69   | 0.57     | 144     |
| 15           | 0.50      | 0.41   | 0.45     | 39      |
| 16           | 0.40      | 0.09   | 0.14     | 23      |
| 17           | 0.38      | 0.43   | 0.40     | 144     |
| 18           | 0.27      | 0.09   | 0.13     | 34      |
| 19           | 0.20      | 0.04   | 0.07     | 24      |
| 20           | 0.36      | 0.38   | 0.37     | 104     |
| 21           | 1.00      | 0.20   | 0.33     | 15      |
| 22           | 0.45      | 0.69   | 0.54     | 45      |
| 23           | 0.50      | 0.17   | 0.25     | 24      |
| 24           | 0.65      | 0.61   | 0.63     | 116     |
| 25           | 0.64      | 0.47   | 0.54     | 15      |
| 26           | 0.52      | 0.32   | 0.40     | 37      |
| 27           | 0.76      | 0.79   | 0.78     | 155     |
| 28           | 0.90      | 0.58   | 0.70     | 33      |
| 29           | 0.75      | 0.75   | 0.75     | 83      |
| 30           | 0.74      | 0.54   | 0.63     | 59      |
| 31           | 0.49      | 0.56   | 0.52     | 39      |
| 32           | 0.77      | 0.22   | 0.34     | 46      |
| 33           | 0.00      | 0.00   | 0.00     | 13      |
| 34           | 0.26      | 0.50   | 0.34     | 20      |
| 35           | 0.00      | 0.00   | 0.00     | 14      |
| 36           | 0.00      | 0.00   | 0.00     | 16      |
| accuracy     |           |        | 0.57     | 2297    |
| macro avg    | 0.55      | 0.42   | 0.44     | 2297    |
| weighted avg | 0.57      | 0.57   | 0.54     | 2297    |
|              |           |        |          |         |

# XGB Classifier

| 1 0.51 0.40 0.45 47 2 0.68 0.45 0.54 29 3 0.58 0.48 0.53 62 4 0.32 0.52 0.39 125 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 24 0.60 0.50 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.56 1.64 25 0.42 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297  |          | precision | recall | f1-score | support |
|--|----------|-----------|--------|----------|---------|
| 1 0.51 0.40 0.45 47 2 0.68 0.45 0.54 29 3 0.58 0.48 0.53 62 4 0.32 0.52 0.39 125 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 24 0.60 0.50 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.56 1.64 25 0.42 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297  | 0        | 0.65      | 0.88   | 0.75     | 203     |
| 2 0.68 0.45 0.54 29 3 0.58 0.48 0.53 62 4 0.32 0.52 0.39 125 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.56 116 25 0.42 0.53 0.67 15 26 0.50 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.79 0.79 0.79 155 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297  |          |           |        |          |         |
| 3 0.58 0.48 0.53 62 4 0.32 0.52 0.39 125 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 4 0.60 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.56 0.58 116 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16 | 2        |           |        |          |         |
| 4 0.32 0.52 0.39 125 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.55 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.55 0.29 13 0.40 0.15 0.22 13 0.40 0.15 0.22 13 0.40 0.15 0.22 13 0.40 0.57 2297 0.60 0.50 0.50 0.64 16   |          |           |        |          |         |
| 5 0.45 0.26 0.33 34 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.56 116 25 0.42 0.53 0.67 15 26 0.50 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.79 0.79 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297   |          |           |        |          |         |
| 6 0.65 0.65 0.65 127 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 4 0.60 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.57 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.32 0.50 0.39 20 35 0.33 0.21 0.26 46 36 0.89 0.50 0.64 16  | 5        | 0.45      |        | 0.33     | 34      |
| 7 0.45 0.33 0.38 51 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297  |          |           |        |          |         |
| 8 0.90 0.69 0.78 13 9 0.96 0.95 0.96 80 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.67 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297  |          |           |        |          |         |
| 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297   | 8        | 0.90      | 0.69   |          | 13      |
| 10 0.48 0.60 0.53 70 11 0.65 0.45 0.53 29 12 0.62 0.64 0.63 158 13 0.25 0.07 0.11 27 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.60 0.57 2297 macro avg 0.54 0.47 0.49 2297   | 9        | 0.96      | 0.95   | 0.96     | 80      |
| 12   |          |           |        |          | 70      |
| 13   | 11       | 0.65      | 0.45   | 0.53     | 29      |
| 14 0.49 0.66 0.56 144 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 accuracy 0.57 2297 macro avg 0.54 0.47 0.49 2297   | 12       | 0.62      | 0.64   | 0.63     | 158     |
| 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.49 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 4 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 1.79 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  |          |           |        |          |         |
| 15 0.47 0.44 0.45 39 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 1.79 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 accuracy 0.57 2297 macro avg 0.54 0.47 0.49 2297  | 14       | 0.49      | 0.66   | 0.56     | 144     |
| 16 0.15 0.09 0.11 23 17 0.49 0.36 0.41 144 18 0.08 0.03 0.04 34 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 0.60 0.56 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 0.89 0.50 0.64 16  | 15       | 0.47      | 0.44   |          | 39      |
| 18   | 16       | 0.15      | 0.09   | 0.11     | 23      |
| 19 0.33 0.17 0.22 24 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  | 17       | 0.49      | 0.36   | 0.41     | 144     |
| 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16   | 18       | 0.08      | 0.03   | 0.04     | 34      |
| 20 0.41 0.44 0.43 104 21 0.89 0.53 0.67 15 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16   | 19       | 0.33      | 0.17   | 0.22     | 24      |
| 22 0.45 0.67 0.54 45 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.38 0.21 0.26 14 36 0.89 0.50 0.64 16  |          |           |        |          |         |
| 23 0.33 0.17 0.22 24 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16   | 21       | 0.89      | 0.53   | 0.67     | 15      |
| 24 0.60 0.56 0.58 116 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 1.55 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297  | 22       | 0.45      | 0.67   | 0.54     | 45      |
| 25 0.42 0.53 0.47 15 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 155 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16   | 23       | 0.33      | 0.17   | 0.22     | 24      |
| 26 0.50 0.27 0.35 37 27 0.79 0.79 0.79 1.55 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297   | 24       | 0.60      | 0.56   | 0.58     | 116     |
| 27 0.79 0.79 0.79 1.55 28 0.96 0.73 0.83 33 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297  | 25       | 0.42      | 0.53   | 0.47     | 15      |
| 28   | 26       | 0.50      | 0.27   | 0.35     | 37      |
| 29 0.79 0.80 0.79 83 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297  | 27       | 0.79      | 0.79   | 0.79     | 155     |
| 30 0.74 0.63 0.68 59 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297   | 28       | 0.96      | 0.73   | 0.83     | 33      |
| 31 0.44 0.51 0.48 39 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297  | 29       | 0.79      | 0.80   | 0.79     | 83      |
| 32 0.38 0.20 0.26 46 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297   | 30       | 0.74      | 0.63   | 0.68     | 59      |
| 33 0.40 0.15 0.22 13 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16  accuracy macro avg 0.54 0.47 0.49 2297  | 31       | 0.44      | 0.51   | 0.48     | 39      |
| 34 0.32 0.50 0.39 20 35 0.33 0.21 0.26 14 36 0.89 0.50 0.64 16 accuracy 0.54 0.47 0.49 2297  | 32       | 0.38      | 0.20   | 0.26     | 46      |
| 35 0.33 0.21 0.26 14<br>36 0.89 0.50 0.64 16<br>accuracy 0.57 2297<br>macro avg 0.54 0.47 0.49 2297  | 33       | 0.40      | 0.15   | 0.22     | 13      |
| 36 0.89 0.50 0.64 16  accuracy 0.57 2297  macro avy 0.54 0.47 0.49 2297  | 34       | 0.32      | 0.50   | 0.39     | 20      |
| accuracy 0.57 2297<br>macro avg 0.54 0.47 0.49 2297  | 35       | 0.33      | 0.21   | 0.26     | 14      |
| macro avg 0.54 0.47 0.49 2297  | 36       | 0.89      | 0.50   | 0.64     | 16      |
| macro avg 0.54 0.47 0.49 2297  | accuracy |           |        | 0.57     | 2297    |
|  |          | 0.54      | 0.47   |          |         |
|  |          |           |        |          |         |

# **Multinomial Naive Bayes**

|              | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.14      | 1.00   | 0.25     | 203     |
| 1            | 0.00      | 0.00   | 0.00     | 47      |
| 2            | 0.00      | 0.00   | 0.00     | 29      |
| 3            | 1.00      | 0.02   | 0.03     | 62      |
| 4            | 0.60      | 0.02   | 0.05     | 125     |
| 5            | 0.00      | 0.00   | 0.00     | 34      |
| 6            | 0.72      | 0.49   | 0.58     | 127     |
| 7            | 0.00      | 0.00   | 0.00     | 51      |
| 8            | 0.00      | 0.00   | 0.00     | 13      |
| 9            | 1.00      | 0.79   | 0.88     | 80      |
| 10           | 0.45      | 0.56   | 0.50     | 70      |
| 11           | 0.00      | 0.00   | 0.00     | 29      |
| 12           | 0.57      | 0.44   | 0.50     | 158     |
| 13           | 0.00      | 0.00   | 0.00     | 27      |
| 14           | 0.32      | 0.69   | 0.44     | 144     |
| 15           | 0.00      | 0.00   | 0.00     | 39      |
| 16           | 0.00      | 0.00   | 0.00     | 23      |
| 17           | 0.32      | 0.17   | 0.22     | 144     |
| 18           | 0.00      | 0.00   | 0.00     | 34      |
| 19           | 0.00      | 0.00   | 0.00     | 24      |
| 20           | 0.33      | 0.01   | 0.02     | 104     |
| 21           | 0.00      | 0.00   | 0.00     | 15      |
| 22           | 0.29      | 0.04   | 0.08     | 45      |
| 23           | 0.00      | 0.00   | 0.00     | 24      |
| 24           | 0.86      | 0.05   | 0.10     | 116     |
| 25           | 0.00      | 0.00   | 0.00     | 15      |
| 26           | 0.00      | 0.00   | 0.00     | 37      |
| 27           | 0.70      | 0.59   | 0.64     | 155     |
| 28           | 0.00      | 0.00   | 0.00     | 33      |
| 29           | 1.00      | 0.10   | 0.18     | 83      |
| 30           | 0.00      | 0.00   | 0.00     | 59      |
| 31           | 0.00      | 0.00   | 0.00     | 39      |
| 32           | 0.00      | 0.00   | 0.00     | 46      |
| 33           | 0.00      | 0.00   | 0.00     | 13      |
| 34           | 0.00      | 0.00   | 0.00     | 20      |
| 35           | 0.00      | 0.00   | 0.00     | 14      |
| 36           | 0.00      | 0.00   | 0.00     | 16      |
| accuracy     |           |        | 0.29     | 2297    |
| macro avo    | 0.22      | 0.13   | 0.12     | 2297    |
| weighted avg | 0.39      | 0.29   | 0.24     | 2297    |
|              |           |        |          |         |

# K-Nearest Neighbors(KNN)

| 1 0.07 0.09 0.08 47 2 0.29 0.34 0.32 29 3 0.27 0.44 0.34 62 4 0.09 0.70 0.16 125 5 0.33 0.03 0.05 34 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.7 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 0.05 27 0.75 0.37 0.50 155 28 1.00 0.22 0.33 159 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 33 34 0.17 0.05 0.31 59 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.84 0.27 0.93 155 28 1.00 0.10 0.10 15 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 36 0.00 0.00 0.00 16 |              | precision | recall | f1-score | support |
|--|--------------|-----------|--------|----------|---------|
| 1 0.07 0.09 0.08 47 2 0.29 0.34 0.32 29 3 0.27 0.44 0.34 62 4 0.09 0.70 0.16 125 5 0.33 0.03 0.05 34 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.7 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 0.05 27 0.75 0.37 0.50 155 28 1.00 0.22 0.33 159 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 33 34 0.17 0.05 0.31 59 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.84 0.27 0.93 155 28 1.00 0.10 0.10 15 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 36 0.00 0.00 0.00 16 | 0            | 0.48      | 0.81   | 0.60     | 203     |
| 2 0.29 0.34 0.32 29 3 0.27 0.44 0.34 62 4 0.09 0.70 0.16 125 5 0.33 0.03 0.05 34 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 12 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.05 144 18 0.38 0.13 0.19 39 16 0.33 0.04 0.15 144 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 14 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 12 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 24 27 0.75 0.37 0.50 155 28 1.00 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 0.00 16   | 1            |           |        |          |         |
| 3 0.27 0.44 0.34 62 4 0.09 0.70 0.16 125 5 0.33 0.03 0.05 34 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 24 27 0.75 0.37 0.50 155 28 1.00 0.20 0.31 59 31 0.30 0.61 0.35 33 30 0.63 0.20 0.31 59 31 0.30 0.61 0.35 33 30 0.63 0.20 0.31 59 31 0.30 0.61 0.35 30 34 0.17 0.05 0.36 33 36 0.63 0.20 0.31 59 31 0.30 0.60 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.00 0.00 13 36 0.00 0.00 0.00 13 37 0.50 0.31 59 38 0.00 0.00 0.00 0.00 13 38 0.17 0.05 0.08 20 39 0.64 0.08 20 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 20 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14  | 2            | 0.29      | 0.34   | 0.32     | 29      |
| 5 0.33 0.03 0.05 34 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 0.36 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 17 0.26 0.10 0.15 144 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 14 18 0.38 0.99 0.14 34 19 0.00 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.84 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 46 33 0.00 0.00 0.00 0.00 16  |              |           | 0.44   |          | 62      |
| 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.84 0.88 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.12 15  | 4            | 0.09      | 0.70   | 0.16     | 125     |
| 6 0.69 0.34 0.46 127 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.84 0.88 46 33 0.00 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.12 15  | 5            | 0.33      | 0.03   | 0.05     | 34      |
| 7 0.52 0.22 0.31 51 8 0.78 0.54 0.64 13 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 24 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.00 0.00 16  |              | 0.69      | 0.34   | 0.46     | 127     |
| 9 0.95 0.90 0.92 80 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.32 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 24 27 0.75 0.37 0.50 155 28 1.00 0.20 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.61 0.24 39 32 0.29 0.84 39 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.00 0.00 16   | 7            | 0.52      | 0.22   | 0.31     | 51      |
| 10 0.36 0.36 0.36 70 11 1.00 0.24 0.39 29 12 0.49 0.22 0.31 158 13 0.11 0.04 0.06 27 14 0.44 0.22 0.30 144 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.03 159 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13   | 8            | 0.78      | 0.54   | 0.64     | 13      |
| 10   | 9            | 0.95      | 0.90   | 0.92     | 80      |
| 12   | 10           | 0.36      | 0.36   | 0.36     | 70      |
| 13   | 11           | 1.00      | 0.24   | 0.39     | 29      |
| 14   | 12           | 0.49      | 0.22   | 0.31     | 158     |
| 15 0.38 0.13 0.19 39 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.44 0.88 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  | 13           | 0.11      | 0.04   | 0.06     | 27      |
| 16 0.33 0.04 0.08 23 17 0.26 0.10 0.15 144 18 0.38 0.09 0.14 34 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.44 0.27 45 33 0.00 0.00 0.00 30 31 59 31 0.30 0.21 0.24 39 32 0.29 0.44 0.27 39 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 36 0.00 0.00 0.00 16   | 14           | 0.44      | 0.22   | 0.30     | 144     |
| 17   | 15           | 0.38      | 0.13   | 0.19     | 39      |
| 18   | 16           | 0.33      | 0.04   | 0.08     | 23      |
| 19 0.00 0.00 0.00 24 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 0.00 16   | 17           | 0.26      | 0.10   | 0.15     | 144     |
| 20 0.27 0.09 0.13 104 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 38 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16   | 18           | 0.38      | 0.09   | 0.14     | 34      |
| 21 1.00 0.07 0.12 15 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  | 19           | 0.00      | 0.00   | 0.00     | 24      |
| 22 0.29 0.24 0.27 45 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 16 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16   | 20           | 0.27      | 0.09   | 0.13     | 104     |
| 23 0.00 0.00 0.00 24 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16   | 21           | 1.00      | 0.07   | 0.12     | 15      |
| 24 0.48 0.09 0.16 116 25 0.32 0.67 0.43 15 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16   | 22           | 0.29      | 0.24   | 0.27     | 45      |
| 25   | 23           | 0.00      | 0.00   | 0.00     | 24      |
| 26 0.00 0.00 0.00 37 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy macro avg 0.42 0.22 0.24 2297  | 24           | 0.48      | 0.09   | 0.16     | 116     |
| 27 0.75 0.37 0.50 155 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy macro avg 0.42 0.22 0.24 2297   | 25           | 0.32      | 0.67   | 0.43     | 15      |
| 28 1.00 0.12 0.22 33 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy macro avg 0.42 0.22 0.24 2297  | 26           | 0.00      | 0.00   | 0.00     | 37      |
| 29 0.86 0.23 0.36 83 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy macro avg 0.42 0.22 0.24 2297  | 27           | 0.75      | 0.37   | 0.50     | 155     |
| 30 0.63 0.20 0.31 59 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy 0.42 0.22 0.24 2297   | 28           | 1.00      | 0.12   | 0.22     | 33      |
| 31 0.30 0.21 0.24 39 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy 0.30 2297 macro avg 0.42 0.22 0.24 2297  | 29           | 0.86      | 0.23   | 0.36     | 83      |
| 32 0.29 0.04 0.08 46 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy macro avg 0.42 0.22 0.24 2297   | 30           | 0.63      | 0.20   | 0.31     | 59      |
| 33 0.00 0.00 0.00 13 34 0.17 0.05 0.08 20 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16  accuracy 0.30 2297 macro avg 0.42 0.22 0.24 2297  |              |           | 0.21   | 0.24     | 39      |
| 34 0.17 0.05 0.08 20<br>35 1.00 0.07 0.13 14<br>36 0.00 0.00 0.00 16<br>accuracy 0.30 2297<br>macro avg 0.42 0.22 0.24 2297  | 32           |           | 0.04   | 0.08     | 46      |
| 35 1.00 0.07 0.13 14 36 0.00 0.00 0.00 16 accuracy 0.30 2297 macro avg 0.42 0.22 0.24 2297   | 33           |           |        | 0.00     | 13      |
| 36 0.00 0.00 0.00 16  accuracy 0.30 2297 macro avg 0.42 0.22 0.24 2297   |              |           |        |          | 20      |
| accuracy 0.30 2297<br>macro avg 0.42 0.22 0.24 2297  |              | 1.00      | 0.07   |          |         |
| macro avg 0.42 0.22 0.24 2297  | 36           | 0.00      | 0.00   | 0.00     | 16      |
|  | accuracy     |           |        | 0.30     | 2297    |
| weighted avg 0.45 0.30 0.30 2297   | macro avg    | 0.42      | 0.22   | 0.24     | 2297    |
|  | weighted avg | 0.45      | 0.30   | 0.30     | 2297    |

# Random Forest Classifier

|              | precision | recall       | f1-score     | support |
|--------------|-----------|--------------|--------------|---------|
| 0            | 0.09      | 1.00         | 0.17         | 203     |
| 1            | 0.00      | 0.00         | 0.00         | 47      |
| 2            | 0.00      | 0.00         | 0.00         | 29      |
| 3            | 0.00      | 0.00         | 0.00         | 62      |
| 4            | 0.00      | 0.00         | 0.00         | 125     |
| 5            | 0.00      | 0.00         | 0.00         | 34      |
| 6            | 1.00      | 0.07         | 0.13         | 127     |
| 7            | 0.00      | 0.00         | 0.00         | 51      |
| 8            | 0.00      | 0.00         | 0.00         | 13      |
| 9            | 1.00      | 0.31         | 0.48         | 80      |
| 10           | 0.60      | 0.09         | 0.15         | 70      |
| 11           | 0.00      | 0.00         | 0.00         | 29      |
| 12           | 0.00      | 0.00         | 0.00         | 158     |
| 13           | 0.00      | 0.00         | 0.00         | 27      |
| 14           | 0.38      | 0.02         | 0.04         | 144     |
| 15           | 0.00      | 0.00         | 0.00         | 39      |
| 16           | 0.00      | 0.00         | 0.00         | 23      |
| 17           | 0.00      | 0.00         | 0.00         | 144     |
| 18           | 0.00      | 0.00         | 0.00         | 34      |
| 19           | 0.00      | 0.00         | 0.00         | 24      |
| 20           | 0.00      | 0.00         | 0.00         | 104     |
| 21           | 0.00      | 0.00         | 0.00         | 15      |
| 22           | 0.00      | 0.00         | 0.00         | 45      |
| 23           | 0.00      | 0.00         | 0.00         | 24      |
| 24           | 0.00      | 0.00         | 0.00         | 116     |
| 25           | 0.00      | 0.00         | 0.00         | 15      |
| 26           | 0.00      | 0.00         | 0.00         | 37      |
| 27           | 0.92      | 0.08         | 0.14         | 155     |
| 28           | 0.00      | 0.00         | 0.00         | 33      |
| 29           | 0.00      | 0.00         | 0.00         | 83      |
| 30           | 0.00      | 0.00         | 0.00         | 59      |
| 31           | 0.00      | 0.00         | 0.00         | 39      |
| 32           | 0.00      | 0.00         | 0.00         | 46      |
| 33           | 0.00      | 0.00         | 0.00         | 13      |
| 34           | 0.00      | 0.00         | 0.00         | 20      |
| 35           | 0.00      | 0.00         | 0.00         | 14      |
| 36           | 0.00      | 0.00         | 0.00         | 16      |
| 200112301    |           |              | 0.11         | 2207    |
| accuracy     | 0 11      | 0.04         |              | 2297    |
| macro avg    | 0.11      | 0.04<br>0.11 | 0.03<br>0.06 | 2297    |
| weighted avg | 0.20      | 0.11         | 0.00         | 2297    |

# Accuracy/F1 Score

| Logistic Regression      | 0.54 |
|--------------------------|------|
| XGB Classifier           | 0.57 |
| Multinomial Naive Bayes  | 0.29 |
| K-Nearest Neighbors(KNN) | 0.30 |
| Random Forest Classifier | 0.1  |

# Thank you