



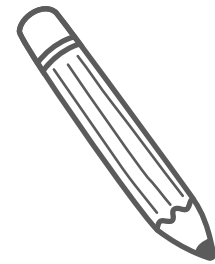
CS 448 - Natural Language Processing

Project Presentation

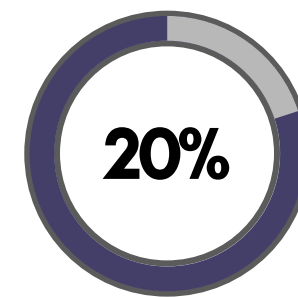
What is Dyslexia?



Dyslexia is a Learning Difficulty.



Makes it challenging to read & write, which makes it difficult to learn.



Impacts 20% people globally.



Easier to help with an early diagnosis.

How it Manifests in Writing?

<i>Error Type</i>	<i>Example</i>	<i>Sources & Frequency</i>
Add Final 'e'	'toll' spelled 'tolle'	Short vowel: 0.04 Long vowel: 0.20 (Bourassa, et al., 2003) (incl. non-final 'e') 32/1377 (Yannakoudakis, et al., 1983)
Delete Final 'e'	'gone' spelled 'gon'	(incl. non-final 'e') 72/1377 (Yannakoudakis, et al., 1983)
Doubling of Consonant	'pine' spelled 'pinne'	0.45 (Bourassa, et al., 2003) 96/1377 (Yannakoudakis, et al., 1983)
Singling of Consonant	'allow' spelled 'alow'	227/1377 (Yannakoudakis, et al., 1983)
One-to-Many Phoneme Mapping	'photo' spelled 'foto'	Vowel (incl. phonetic errors): 850/1377 Consonant (incl. doubling): 280/1377
Phonetic error	'bank' spelled 'pank'	0.31 (Finucci, et al., 1983)
Silent Consonant elided	'bomb' spelled 'bom'	35/1377 (Yannakoudakis, et al., 1983)
Syllable	'library' spelled 'libry'	Elision: 8/1377 (Yannakoudakis, et al., 1983)
Typing Errors	Very variable	1% clearly identifiable (Yannakoudakis, et al., 1983)
Unstressed Vowel	'record' spelled 'recurd'	Omission of ə: 0.17 (Bourassa, et al., 2003)
Vowel Cluster	'dream' spelled 'drim'	413/1377 (Yannakoudakis, et al., 1983)
Consonant Cluster	'create' spelled 'reate'	0.26 (Bruck, et al., 1990)

2 features that we added on top of these were **vocabulary length** and **no. of spelling errors**

A large yellow arc is positioned on the left side of the slide, extending from the bottom towards the middle.

What Are We Trying To Solve?

Can we train a model on the aforementioned features to classify Dyslexic writing samples?

Data Collection


Instructions

Write the 7 days of the week and the 12 months.

days of the week	months

Instructions

Write a few lines about the picture. It can be anything!



The beautiful ducks swam in the big blue lake. Their feathers were very pretty and quite often the ducks would jump out of the water.

From our self-curated booklet, based on the tasks taken from Ann Arbor's Skillset B, we collected data from Dyslexic and Neurotypical children (ages 5-10 years).

Data Set

NT - 10
samples

LD - 13
samples

Monday	January
Tuesday	February
ganvare, fadre, March,	

swarm
doctor
lawyer
children
babies
grapefruit
squirm
birthday
grand mother
holiday

NT - 26
samples

swarm
doctor
lawyer
children
babies
grapefruit
squirm
birthday
grand mother
holiday

LD - 39
samples

The beautiful ducks swam
in the big blue lake. Their
feathers were very pretty
and quite often the ducks
would jump out of the
water.

LD - 17
samples

The bird is squaking happily on a stick. The fish
is swimming happily in its water tank. The cat's
sleeping in its comfy bed. The dog looks
excited to play with his dog bone and to eat
his dog food.

NT - 16
samples

Data Set

LD

Monday

Tuesday

January

February

NT

ganvare,fadre, March,

swarm

doctor

lawyer

children

babies

babies

grapefruit

squirrel

birthday

grand mother

holiday

swarm

doctor

lawyer

children

babies

babies

grapefruit

squirrel

birthday

grand mother

holiday

LD

NT

LD

The beautiful ducks sawm
inth big blue lake. Thir
feather's were uery pretty
and quite often the ducks
would jump out of the
water.

NT

The beautiful ducks swam in the
big blue lake. Their feathers were very
pretty and quite often the ducks would
jump out of the water.

Scanned
Images

3-ld-8 (1) - Notepad

File Edit Format View Help

The beautiful ducks sawm inth big blue lake. Thir feathers were uery pretty and quite often the ducks would jump out of the water.

30-nt-8 (1) - Notepad

File Edit Format View Help

The beautiful ducks swam in the big blue lake. Their feathers were very pretty and quite often the ducks would jump out of the water.

Manually
Transcribed
Copy

Data Set

```
def scoring(df):  
  
    sentence_length = list()  
    PHE = [] #phonetic error [6]  
    UNV = [] #unstressed vowel [10]  
    AEE = [] #added extra e [1]  
    DOC = [] #doubling of consonants [3]  
  
    for ind, row in df.iterrows():  
        phe, unv, aee, doc = 0, 0, 0, 0  
        words = row['sentence'].translate(str.maketrans('', '', string.punctuation)).lower()  
        words = words.split()  
  
        #length of sentence  
        if len(words) == 25: sentence_length.append(0)  
        else: sentence_length.append(abs(len(words)-25))  
  
        for word in words:  
            src = ''  
            tgt = word  
            min_dist = 999  
  
            # edit distance to find the closest word if there are spelling errors  
            for x in word_dict_key:  
                dist = ed(word, x)  
                if dist < min_dist:  
                    min_dist = dist  
                    src = x  
            if min_dist > 0: # there is an error  
                a, b, c, d = sim_characters(src, tgt)  
                phe += a  
                unv += b  
                aee += c  
                doc += d
```

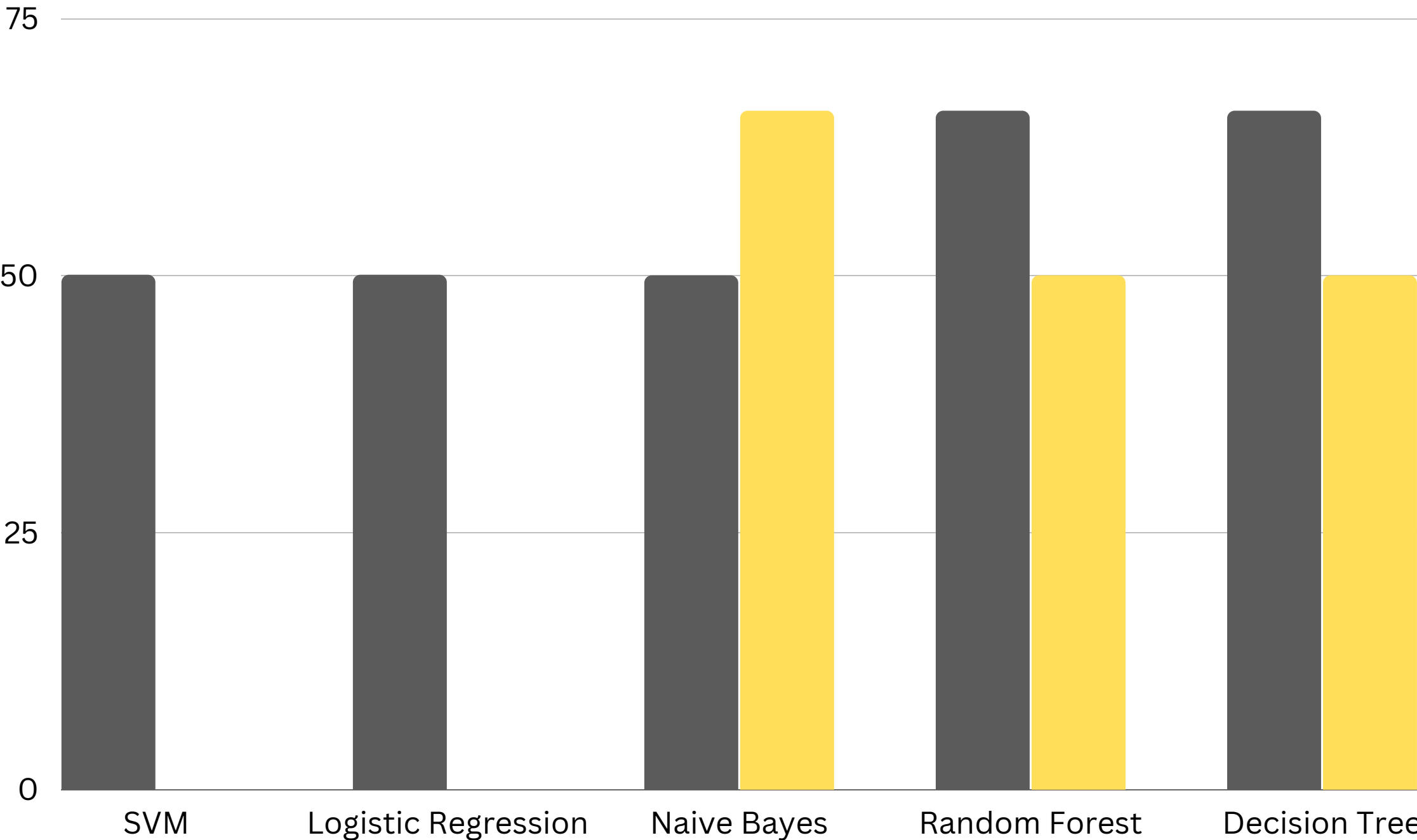
```
data.csv  
1 id,sentence,sentence length,phonetic error,unstressed vowel,adding extra e,doubling consonant  
2 sample,Tbe beautofful dacks swame in the big blue lake. Their feathers were very pretty and quite often the ducks would jump out of the water.,0,1,0,1,1  
3 30_1,The beautiful ducks swam in the big blue lake. Their feathers were very pretty and quite often the ducks would jump out of the water.,0,0,0,0,0  
4 30_5,The fish is not happy because it is in a small fish bowl not in an ocean and it's place is in ocean.,2,0,2,0,0  
5 32_3,cat's are predetors they drinkmilk and is afraid of water. Just like cats dogs are predetors except that they drinkwater. There are many types of  
parrots talking parrots is an example ofaparrot,7,0,0,1,0  
6 33_3,The dog is happy.The kitten is sleeping. The fish is in the bowl. The parrot has a long beak. There is a house at the corner. There are footprints of  
the dog near the parrot,10,4,0,0,1  
7 34-nt_1,The dog is eager to eat Food. The cat is in a deep sleep. The bird is singing happily. The fish is swimming in the water. The dog looks excited to  
chew on the bone and is wagging his tail.,15,2,1,0,0  
8 35-nt_3,The bird is squaking happily on a stick. The fish is swimming happily it's water tank. The cat's sleeping in it's comfy bed. The dog looks exited  
to play with his dog bone and to eat his dog food.,14,3,2,0,0
```

CSV for extracted features based on the guide

```
results.csv  
1 Filename,Vocabulary Length,Complexity,Spelling Errors  
2 D:\OneDrive - Habib University\Kaavish\nlp-spelling-dysign\data_transcribed\1-MAR001-1.txt,21,2,  
3 D:\OneDrive - Habib University\Kaavish\nlp-spelling-dysign\data_transcribed\12-MAR011_1.txt,21,2,  
4 D:\OneDrive - Habib University\Kaavish\nlp-spelling-dysign\data_transcribed\13-MAR012_1.txt,20,2,  
5 D:\OneDrive - Habib University\Kaavish\nlp-spelling-dysign\data_transcribed\26-DIY001_2.txt,21,2,  
6 D:\OneDrive - Habib University\Kaavish\nlp-spelling-dysign\data_transcribed\27-DIY002_1.txt,20,2,
```

CSV for recording spelling errors & word complexity in each sample

Model - Spelling Errors



SVM Classifier:
Accuracy: 0.5
Precision: 0.0
Recall: 0.0
F1 Score: 0.0

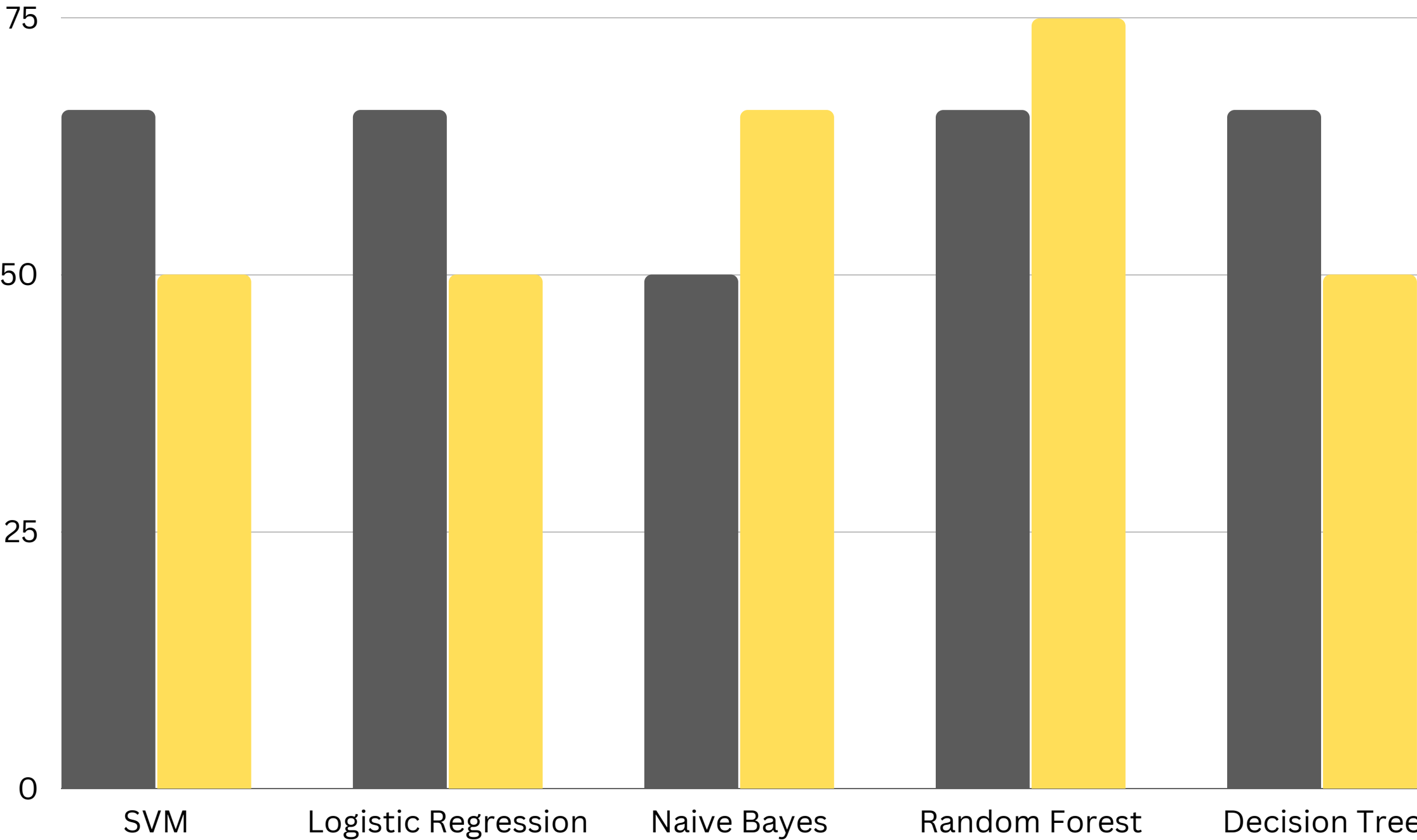
Logistic Regression Classifier:
Accuracy: 0.5
Precision: 0.0
Recall: 0.0
F1 Score: 0.0

Naive Bayes Classifier:
Accuracy: 0.5
Precision: 0.5
Recall: 1.0
F1 Score: 0.6666666666666666

Random Forest Classifier:
Accuracy: 0.6666666666666666
Precision: 1.0
Recall: 0.3333333333333333
F1 Score: 0.5

Decision Tree Classifier:
Accuracy: 0.6666666666666666
Precision: 1.0
Recall: 0.3333333333333333
F1 Score: 0.5

Model - Vocabulary & Word Complexity



SVM Classifier:
Accuracy: 0.6666666666666666
Precision: 1.0
Recall: 0.3333333333333333
F1 Score: 0.5

Logistic Regression Classifier:
Accuracy: 0.6666666666666666
Precision: 1.0
Recall: 0.3333333333333333
F1 Score: 0.5

Naive Bayes Classifier:
Accuracy: 0.5
Precision: 0.5
Recall: 1.0
F1 Score: 0.6666666666666666

Random Forest Classifier:
Accuracy: 0.6666666666666666
Precision: 0.6
Recall: 1.0
F1 Score: 0.7499999999999999

Decision Tree Classifier:
Accuracy: 0.6666666666666666
Precision: 1.0
Recall: 0.3333333333333333
F1 Score: 0.5

Results & Discussion

- For character-level errors in words, Naive Bayes Classifier performed the best out of all 5, however, the accuracy was about 66%.
- For vocabulary and word complexity, random forest worked the best with 66% accuracy and almost 75% F1 score.

Limitations & Challenges

- Small data set due to manual data collection
- Difficulty in OCR due to the varying and sometimes illegible handwriting.

Future Work

- The curated booklets can be used for more data collection on a larger scale.
- Better OCR techniques or scoring system for illegible characters.
- Better model training.