Predicting Authors of Bible Passages and Bible Passages with ML

November 21, 2019

Predicting Bible Passage and Bible Authors Using Machine Learning

- Dataset : KJV

In [1]:
```
# Load EDA Pkgs
import pandas as pd
```

In [2]:
```
# Load ML Pkgs
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
```

In [3]:
```
# Load Dataset
df = pd.read_csv("kjvdata.csv")
```

In [4]:
```
df.head()
```

Out[4]:

	Unnamed: 0	id	book	chapter	verse	text
0	0	1001001	Genesis	1	1	In the beginning God created the heaven and th...
1	1	1001002	Genesis	1	2	And the earth was without form, and void; and ...
2	2	1001003	Genesis	1	3	And God said, Let there be light: and there wa...
3	3	1001004	Genesis	1	4	And God saw the light, that it was good: and G...
4	4	1001005	Genesis	1	5	And God called the light Day, and the darkness...

In [5]:
```
# Columns
df.columns
```

Out[5]:
```
Index(['Unnamed: 0', 'id', 'book', 'chapter', 'verse', 'text'], dtype='object')
```

In [6]:
```
df['book'].unique()
```

```
Out[6]: array(['Genesis', 'Exodus', 'Leviticus', 'Numbers', 'Deuteronomy',
               'Joshua', 'Judges', 'Ruth', '1 Samuel (1 Kings)',
               '2 Samuel (2 Kings)', '1 Kings (3 Kings)', '2 Kings (4 Kings)',
               '1 Chronicles', '2 Chronicles', 'Ezra', 'Nehemiah', 'Esther',
               'Job', 'Psalms', 'Proverbs', 'Ecclesiastes',
               'Song of Solomon (Canticles)', 'Isaiah', 'Jeremiah',
               'Lamentations', 'Ezekiel', 'Daniel', 'Hosea', 'Joel', 'Amos',
               'Obadiah', 'Jonah', 'Micah', 'Nahum', 'Habakkuk', 'Zephaniah',
               'Haggai', 'Zechariah', 'Malachi', 'Matthew', 'Mark', 'Luke',
               'John', 'Acts', 'Romans', '1 Corinthians', '2 Corinthians',
               'Galatians', 'Ephesians', 'Philippians', 'Colossians',
               '1 Thessalonians', '2 Thessalonians', '1 Timothy', '2 Timothy',
               'Titus', 'Philemon', 'Hebrews', 'James', '1 Peter', '2 Peter',
               '1 John', '2 John', '3 John', 'Jude', 'Revelation'], dtype=object)
```

```
In [7]: author_list = {"Genesis": "Moses",
        "Exodus": "Moses",
        "Leviticus": "Moses",
        "Numbers": "Moses",
        "Deuteronomy": "Moses",
        "Joshua": "Joshua",
        "Judges": "Samuel, Nathan, Gad",
        "Ruth": "Samuel, Nathan, Gad",
        "1 Samuel (1 Kings)": "Samuel, Nathan, Gad",
        "2 Samuel (2 Kings)": "Samuel, Nathan, Gad",
        "1 Kings (3 Kings)": "Jeremiah",
        "2 Kings (4 Kings)": "Jeremiah",
        "1 Chronicles": "Ezra",
        "2 Chronicles": "Ezra",
        "Ezra": "Ezra",
        "Nehemiah": "Nehemiah, Ezra",
        "Esther": "Mordecai",
        "Job": "Job,Moses",
        "Psalms": "David,Asaph, Ezra, the sons of Korah, Heman, Ethan, Moses",
        "Proverbs": "Solomon ,Agur(30) and Lemuel(31)",
        "Ecclesiastes": "Solomon",
        "Song of Solomon (Canticles)": "Solomon",
        "Isaiah": "Isaiah",
        "Jeremiah": "Jeremiah",
        "Lamentations": "Jeremiah",
        "Ezekiel": "Ezekiel",
        "Daniel": "Daniel",
        "Hosea": "Hosea",
        "Joel": "Joel",
        "Amos": "Amos",
        "Obadiah": "Obadiah",
        "Jonah": "Jonah",
        "Micah": "Micah",
```

```
        "Nahum": "Nahum",
        "Habakkuk": "Habakkuk",
        "Zephaniah": "Zephaniah",
        "Haggai": "Haggai",
        "Zechariah": " Zechariah",
        "Malachi": "Malachi",
        "Matthew": "Matthew",
        "Mark": "John Mark",
        "Luke": "Luke",
        "John": "John, the Apostle",
        "Acts": "Luke",
        "Romans": "Paul",
        "1 Corinthians": "Paul",
        "2 Corinthians": "Paul",
        "Galatians": "Paul",
        "Ephesians": "Paul",
        "Philippians": "Paul",
        "Colossians": "Paul",
        "1 Thessalonians": "Paul",
        "2 Thessalonians": "Paul",
        "1 Timothy": "Paul",
        "2 Timothy": "Paul",
        "Titus": "Paul",
        "Philemon": "Paul",
        "Hebrews": "Paul, Luke, Barnabas, Apollos",
        "James": "James the brother of Jesus and Jude (not the Apostle, brother of John).",
        "1 Peter": "Peter",
        "2 Peter": "Peter",
        "1 John": "John, the Apostle",
        "2 John": "John, the Apostle",
        "3 John": "John, the Apostle",
        "Jude": "Jude, the brother of Jesus",
        "Revelation": "John, the Apostle"}
```

```
In [8]: # Map The Authors to their Books
        df['author'] = df['book'].map(author_list)
```

```
In [9]: df['author'].head()
```

```
Out[9]: 0    Moses
        1    Moses
        2    Moses
        3    Moses
        4    Moses
        Name: author, dtype: object
```

```
In [10]: df.head()
```

```
Out[10]:    Unnamed: 0       id     book  chapter  verse  \
         0           0  1001001  Genesis        1      1
```

```
1           1  1001002  Genesis        1      2
2           2  1001003  Genesis        1      3
3           3  1001004  Genesis        1      4
4           4  1001005  Genesis        1      5

                                                  text author
0  In the beginning God created the heaven and th...  Moses
1  And the earth was without form, and void; and ...  Moses
2  And God said, Let there be light: and there wa...  Moses
3  And God saw the light, that it was good: and G...  Moses
4  And God called the light Day, and the darkness...  Moses
```

```
In [11]: df.columns
```

```
Out[11]: Index(['Unnamed: 0', 'id', 'book', 'chapter', 'verse', 'text', 'author'], dtype='obje
```

```
In [12]: # Label Old and New Testament
         # 23214 = OT
```

```
In [13]: df.shape
```

```
Out[13]: (31103, 7)
```

```
In [14]: df2 = df
```

```
In [15]: df2.iloc[23144]
```

```
Out[15]: Unnamed: 0                                                    23144
         id                                                         39004006
         book                                                        Malachi
         chapter                                                           4
         verse                                                             6
         text          And he shall turn the heart of the fathers to ...
         author                                                      Malachi
         Name: 23144, dtype: object
```

```
In [16]: df2.loc[0:23144,'label'] = 0
```

```
In [17]: df2.loc[23145:,'label'] = 1
```

```
In [18]: df2.head()
```

```
Out[18]:    Unnamed: 0       id     book  chapter  verse  \
         0           0  1001001  Genesis        1      1
         1           1  1001002  Genesis        1      2
         2           2  1001003  Genesis        1      3
         3           3  1001004  Genesis        1      4
         4           4  1001005  Genesis        1      5

                                                  text author  label
```

```
0  In the beginning God created the heaven and th...  Moses    0.0
1  And the earth was without form, and void; and ...  Moses    0.0
2  And God said, Let there be light: and there wa...  Moses    0.0
3  And God saw the light, that it was good: and G...  Moses    0.0
4  And God called the light Day, and the darkness...  Moses    0.0
```

```
In [19]: df2[['id', 'book', 'chapter', 'verse', 'text', 'author','label']].to_csv("KJV_Dataset
```

```
In [20]: #  Features
         Xfeatures = df['text']
         ylabel = df['author']
```

```
In [21]: # Vectorization
         cv = CountVectorizer()
         X = cv.fit_transform(Xfeatures)
```

0.0.1 Save Vectorizer

```
In [22]: # Load Joblib
         import joblib
```

```
In [23]: bible_author_vectorizer = open("bible_author_vectorizer.pkl","wb")
         joblib.dump(cv,bible_author_vectorizer)
```

```
In [24]: bible_author_vectorizer.close()
```

0.0.2 Train Test Split

```
In [25]: X_train,X_test,y_train,y_test  = train_test_split(X,ylabel,test_size=0.33,random_stat
```

```
In [26]: # Model Building
         clf = MultinomialNB()
         clf.fit(X_train,y_train)
```

```
Out[26]: MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

```
In [27]: # Accuracy Score of our r Model
         print("Accuracy of Training",clf.score(X_train,y_train))
```

```
Accuracy of Training 0.6184557800278324
```

```
In [28]: # Accuracy Score of our Model
         print("Accuracy of model",clf.score(X_test,y_test))
```

```
Accuracy of model 0.5213367108339828
```

0.0.3 Predicting A Passage

- Whether therefore ye eat,or drink,or whatsoever ye do,do all to the glory of God by Paul

```
In [29]: sample_verse =["Whether therefore ye eat,or drink,or whatsoever ye do,do all to the gl
```

```
In [30]: # Vectorize Text
         vect = cv.transform(sample_verse).toarray()
```

```
In [31]: # Predict
         clf.predict(vect)
```

```
Out[31]: array(['Paul'], dtype='<U71')
```

Save Model

```
In [32]: # from sklearn.externals import joblib
         import joblib
```

```
In [33]: bible_author_NV_model = open("bible_author_prediction_NV_model_new.pkl","wb")
         joblib.dump(clf,bible_author_NV_model)
```

```
In [34]: bible_author_NV_model.close()
```

Using Logistic Regression

```
In [35]: from sklearn.linear_model import LogisticRegression
         from sklearn.metrics import accuracy_score
```

```
In [36]: logit = LogisticRegression()
         logit.fit(X_train,y_train)
```

```
/usr/local/lib/python3.6/dist-packages/sklearn/linear_model/logistic.py:433: FutureWarning: De
  FutureWarning)
/usr/local/lib/python3.6/dist-packages/sklearn/linear_model/logistic.py:460: FutureWarning: De
  "this warning.", FutureWarning)
```

```
Out[36]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='warn',
                   n_jobs=None, penalty='l2', random_state=None, solver='warn',
                   tol=0.0001, verbose=0, warm_start=False)
```

```
In [37]: print("Model Accuracy:", accuracy_score(y_test,logit.predict(X_test)))
```

```
Model Accuracy: 0.5965510522213562
```

```
In [38]: # For Training
         logit.score(X_train,y_train)
```

```
Out[38]: 0.8937089111761601
```

```
In [39]: # For Test Dataset
         logit.score(X_test,y_test)
```

```
Out[39]: 0.5965510522213562
```

Predicting with Logistic Regression Model

- Using Same Verse

```
In [40]: logit.predict(vect)

Out[40]: array(['Paul'], dtype=object)

In [41]: # Save Our Model
         bible_author_Logit_model = open("bible_author_prediction_Logit_model_new.pkl","wb")
         joblib.dump(logit,bible_author_Logit_model)

In [42]: bible_author_Logit_model.close()
```

Prediction Location (Old Testament or New Testament)

```
In [43]: #  Features
         Xfeatures2 = df['text']
         ylabel2 = df['label']

In [44]: # Vectorization
         cv2 = CountVectorizer()
         X2 = cv2.fit_transform(Xfeatures)

In [45]: bible_passage_vectorizer = open("bible_passage_vectorizer.pkl","wb")
         joblib.dump(cv2,bible_passage_vectorizer)

In [46]: bible_passage_vectorizer.close()

In [47]: X_train2,X_test2,y_train2,y_test2  = train_test_split(X2,ylabel2,test_size=0.33,rando

In [48]: # Model Building
         clf2 = MultinomialNB()
         clf2.fit(X_train2,y_train2)

Out[48]: MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)

In [49]: # Accuracy Score of our Model
         print("Accuracy of model",clf2.score(X_test2,y_test2))

Accuracy of model 0.9158222915042868
```

0.0.4 Predicting A Bible Passage

```
In [50]: # Vectorize Text
         vect2 = cv2.transform(sample_verse).toarray()

In [51]: clf2.predict(vect2)

Out[51]: array([1.])
```

Narative

- 1 represent New Testament

```
In [52]: bible_passage_NV_model = open("bible_passage_prediction_NV_model_new.pkl","wb")
         joblib.dump(clf2,bible_passage_NV_model)

In [53]: bible_passage_NV_model.close()
```

Using LogisticRegression

```
In [54]: logit2 = LogisticRegression()
         logit2.fit(X_train2,y_train2)

/usr/local/lib/python3.6/dist-packages/sklearn/linear_model/logistic.py:433: FutureWarning: De
  FutureWarning)


Out[54]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                   intercept_scaling=1, max_iter=100, multi_class='warn',
                   n_jobs=None, penalty='l2', random_state=None, solver='warn',
                   tol=0.0001, verbose=0, warm_start=False)

In [55]: # For Test Dataset
         logit2.score(X_test2,y_test2)

Out[55]: 0.9250779423226813

In [56]: logit2.predict(vect)

Out[56]: array([1.])

In [57]: bible_passage_Logit_model = open("bible_passage_prediction_Logit_model_new.pkl","wb")
         joblib.dump(logit2,bible_passage_Logit_model)

In [58]: bible_passage_Logit_model.close()
```

0.0.5 Interpeting Model

- Eli5
- Lime

```
In [56]: # Load Pkgs for Model Interpreting
         import eli5

In [57]: # Show Weight For Our Model
         eli5.show_weights(logit)

Out[57]: <IPython.core.display.HTML object>

In [58]: df['author'].unique()
```

```
Out[58]: array(['Moses', 'Joshua', 'Samuel, Nathan, Gad', 'Jeremiah', 'Ezra',
                'Nehemiah, Ezra', 'Mordecai', 'Job,Moses',
                'David,Asaph, Ezra, the sons of Korah, Heman, Ethan, Moses',
                'Solomon ,Agur(30) and Lemuel(31)', 'Solomon', 'Isaiah', 'Ezekiel',
                'Daniel', 'Hosea', 'Joel', 'Amos', 'Obadiah', 'Jonah', 'Micah',
                'Nahum', 'Habakkuk', 'Zephaniah', 'Haggai', ' Zechariah',
                'Malachi', 'Matthew', 'John Mark', 'Luke', 'John, the Apostle',
                'Paul', 'Paul, Luke, Barnabas, Apollos',
                'James the brother of Jesus and Jude (not the Apostle, brother of John).',
                'Peter', 'Jude, the brother of Jesus'], dtype=object)
```

```
In [59]: class_names = ['Moses', 'Joshua', 'Samuel, Nathan, Gad', 'Jeremiah', 'Ezra',
                'Nehemiah, Ezra', 'Mordecai', 'Job,Moses',
                'David,Asaph, Ezra, the sons of Korah, Heman, Ethan, Moses',
                'Solomon ,Agur(30) and Lemuel(31)', 'Solomon', 'Isaiah', 'Ezekiel',
                'Daniel', 'Hosea', 'Joel', 'Amos', 'Obadiah', 'Jonah', 'Micah',
                'Nahum', 'Habakkuk', 'Zephaniah', 'Haggai', ' Zechariah',
                'Malachi', 'Matthew', 'John Mark', 'Luke', 'John, the Apostle',
                'Paul', 'Paul, Luke, Barnabas, Apollos',
                'James the brother of Jesus and Jude (not the Apostle, brother of John).',
                'Peter', 'Jude, the brother of Jesus']
```

```
In [60]: eli5.show_weights(logit,target_names=class_names)
```

```
Out[60]: <IPython.core.display.HTML object>
```

Show Explanation for A Single Prediction

```
In [61]: import numpy as np
```

```
In [62]: # Convert the text to vectors anr reshape as a single sample with reshape(-1,1)
         ex = np.reshape(vect,-1, 1)
```

```
In [63]: eli5.show_prediction(logit,ex,target_names=class_names)
```

```
Out[63]: <IPython.core.display.HTML object>
```

```
In [64]: ### Multiple Sample Prediction
         sample_2 = [" And it came to pass, when men began to multiply on the face of the earth
```

```
In [65]: vect2 = cv.transform(sample_2).toarray()
```

```
In [66]: clf.predict(vect2)
```

```
Out[66]: array(['Moses', 'John, the Apostle',
                'David,Asaph, Ezra, the sons of Korah, Heman, Ethan, Moses'],
               dtype='<U71')
```

Thanks For Your Time

- Jesus Saves @JCharisTech
- By Jesse E.Agbe(JCharis)

```
In [ ]:
```