

```

import os

import pandas as pd

import numpy as np

import textdistance

import re

from collections import Counter


file_path = 'C:/Users/PRATIBHA/Python/auto.py'


words = []

with open(file_path, 'r') as f:
    file_name_data = f.read()
    file_name_data = file_name_data.lower()
    words = re.findall('w+', file_name_data)


# This is our vocabulary
V = set(words)

print("Top ten words in the text are:{words[0:10]}")

print("Total Unique words are {len(V)}.")

word_freq = {}

word_freq = Counter(words)

print(word_freq.most_common()[0:10])

probs = {}

Total = sum(word_freq.values())

for k in word_freq.keys():
    probs[k] = word_freq[k]/Total

def my_autocorrect(input_word):
    input_word = input_word.lower()
    if input_word in V:
        return('Your word seems to be correct')
    else:

```

```
sim = [1-(textdistance.Jaccard(qval=2).distance(v,input_word)) for v in word_freq.keys()]

df = pd.DataFrame.from_dict(probs, orient='index').reset_index()

df = df.rename(columns={'index':'Word', 0:'Prob'})

df['Similarity'] = sim

output = df.sort_values(['Similarity', 'Prob'], ascending=False).head()

return(output)
```