SENTIMENTAL ANALYSIS OF TWITTER COMMENTS

A Project Report

Submitted in fulfillment of the

requirements for the award of the

Certificate for

Internship Program

By

SMARTBRIDGE

Introduction

* Artificial Intelligence

Artificial intelligence (AI) is the branch of computer science which aims to create intelligence machines. Artificial Intelligence (AI) is the study of how computer systems can simulate intelligent processes such as learning, reasoning, and understanding symbolic information in context. It is simply a technique which enables computer to mimic human behavior.

* Neural Networks

A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes for solving Artificial Intelligence (AI) problems.

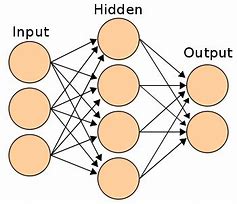


Fig 1.1 A Simple Neural Network

* Convolutional Neural Networks

A Convolutional Neural Network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery.

They have applications in image and video recognition, recommender systems, image classification, medical image analysis, and natural language processing.



Fig 1.2 Convolution Neural Network

Tools/Libraries

* **Languages** : Python
* **Tools/IDE** : Anaconda
* **Libraries :** Keras ,Tokenizer ,Pandas ,numpy

1.Keras:

It is a sub library of TensorFlow

We are using keras and underlying deep learning models to perform data cleaning, normalization and predictions

It offers a higher-level, more intuitive set of abstractions that make it easy to develop deep learning models regardless of the computational backend used

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML.

Designed to enable fast experimentation with deep neural networks, it focuses on being userfriendly, modular, and extensible

**2.Tokenizer**:

In Python tokenization basically refers to splitting up a larger body of text

into smaller lines, words or even creating words for a non-English language. The various tokenization functions in-built into the nltk module itself and can be used in programs as shown below.

Line Tokenization

In the below example we divide a given text into different lines by using the function sent\_tokenize.

import nltk

sentence\_data = "The First sentence is about Python. The Second: about Django. You can learn Python,Django and Data Ananlysis here. "

nltk\_tokens = nltk.sent\_tokenize(sentence\_data)

print (nltk\_tokens)

When we run the above program, we get the following output −

['The First sentence is about Python.', 'The Second: about Django.', 'You can learn Python,Django and Data Ananlysis here.']

Non-English Tokenization

In the below example we tokenize the German text.

import nltk

german\_tokenizer = nltk.data.load('tokenizers/punkt/german.pickle')

german\_tokens=german\_tokenizer.tokenize('Wie geht es Ihnen? Gut, danke.')

print(german\_tokens)

When we run the above program, we get the following output −

['Wie geht es Ihnen?', 'Gut, danke.']

Word Tokenzitaion

We tokenize the words using word\_tokenize function available as part of nltk.

import nltk

word\_data = "It originated from the idea that there are readers who prefer learning new skills from the comforts of their drawing rooms"

nltk\_tokens = nltk.word\_tokenize(word\_data)

print (nltk\_tokens)

Output −

['It', 'originated', 'from', 'the', 'idea', 'that', 'there', 'are', 'readers',

'who', 'prefer', 'learning', 'new', 'skills', 'from', 'the',

'comforts', 'of', 'their', 'drawing', 'rooms']

**3.pandas**

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data.

Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Key Features of Pandas

* Fast and efficient DataFrame object with default and customized indexing.
* Tools for loading data into in-memory data objects from different file formats.
* Data alignment and integrated handling of missing data.
* Reshaping and pivoting of date sets.
* Label-based slicing, indexing and subsetting of large data sets.
* Columns from a data structure can be deleted or inserted.
* Group by data for aggregation and transformations.
* High performance merging and joining of data.
* Time Series functionality.

**4.numpy:**

NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.

**Numeric**, the ancestor of NumPy, was developed by Jim Hugunin. Another package Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created NumPy package by incorporating the features of Numarray into Numeric package. There are many contributors to this open source project.

## Operations using NumPy

Using NumPy, a developer can perform the following operations −

* Mathematical and logical operations on arrays.
* Fourier transforms and routines for shape manipulation.
* Operations related to linear algebra. NumPy has in-built functions for linear algebra and random number generation.

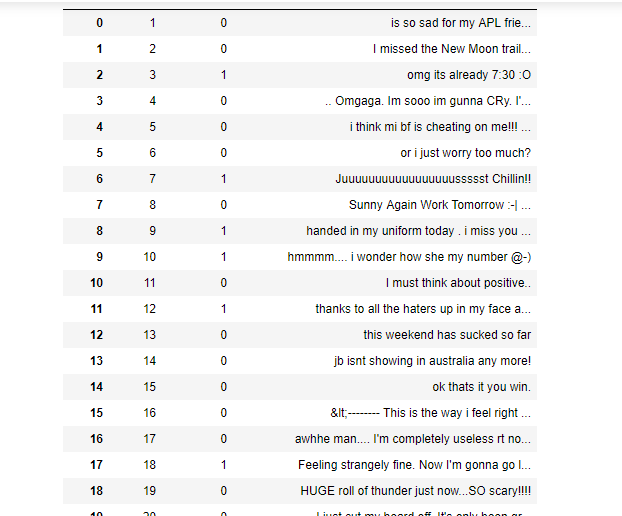
**DataFrame**

**DataFrame** is a 2-dimensional labeled data structure with columns of potentially

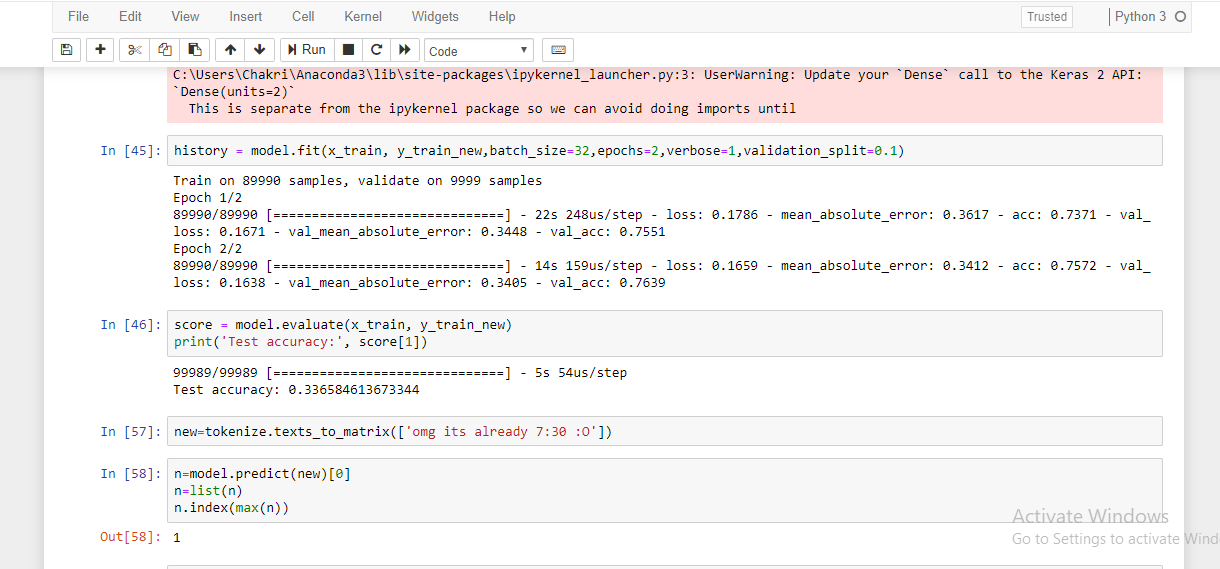
different types. You can think of it like a spreadsheet or SQL table, or a dict of

Series objects. It is generally the most commonly used pandas object.

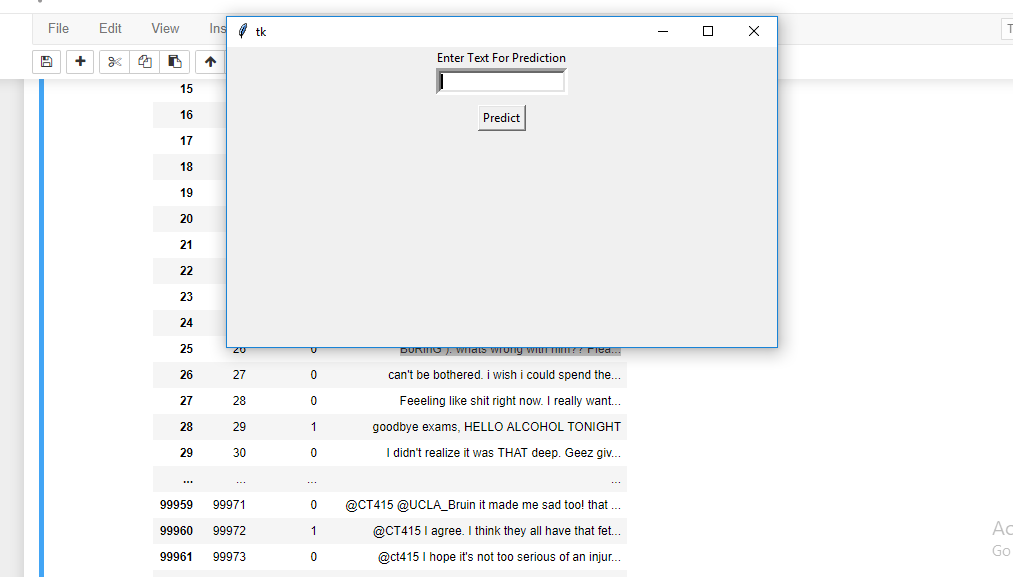
**Data frame used:**



**Accuracy Rate**



Final out predict:



Applications

* Sentiment analysis has many applications and benefits to any business and organizations. It can be used to give a business valuable insights into how people feel about said product, brand or service.
* When applied to social media channels, it can be used to identify spikes in sentiment, thereby allowing you to identify potential product advocates or social media influencers.
* It can be used to identify when potential negative threads are emerging online a business, thereby allowing us to be proactive in dealing with it more quickly.
* Sentiment analysis could also be applied to a corporate network, for example, by applying it to your email server, emails could be monitored for their general “tone