AI MINI PROJECT

Name – Ramya Bardae

PRN- 17070123080

TOPIC – Movie review classification

THEORY-

The data set is taken from Kaggle.com. It basically has 50 thousand IMDB reviews. This data set has to be spilt in such a way that it should have 25 thousand reviews for training and 25 thousand reviews for testing. In this project I am going to predict if the review is a positive or negative review by binary classifications of these movie reviews using TensorFlow and NTKL (natural language tool kit). This project is a sentiment analysis classifier problem.

a) **Sentimental analysis**- It is a technique that determines whether the data give holds a positive or negative result. Mostly, such techniques are used to analyse customer feedback based on which companies can find ways to improve their brand.

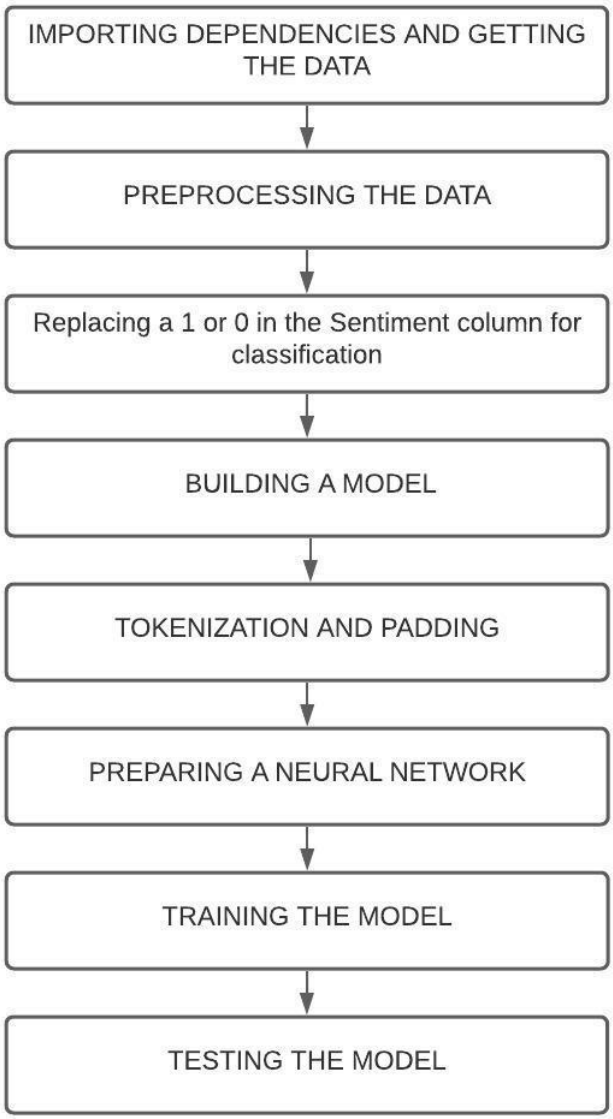
b) **TensorFlow** – It is the largest opens source platform provided by google to help users carry out various machine learning algorithms with an ease. In this project we will use keras which is an open-source library that delivers a Python interface for ANN. Keras behaves as an interface for the TensorFlow library. We will build our neural network using keras.

c) **NLTK (natural language tool kit)**- NLTK is the main platform for constructing python applications to deal with human language data. It affords easy-to-use interfaces to over 50 resources which includes WordNet, alongside a set of textual content processing libraries for semantic reasoning, stemming, classification, tagging, tokenization, parsing, and wrappers for industrial-energy NLP libraries, few of which we will use in the project.
Different methods are used from NLTK library and tensor flow library to help the model predict the correctly

DATASET –

	review
2	One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me.
3	A wonderful little production. The filming technique is very unassuming- very old-time-BBC fashion and gives a comforting, and sometimes discomforting,
4	I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air conditioned theater and watching a light-hearted comedy. The plot i
5	Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents are fighting all the time. This movie is slower than a soa
6	Petter Mattei's "Love in the Time of Money" is a visually stunning film to watch. Mr. Mattei offers us a vivid portrait about human relations. This is a movie that seems t
7	Probably my all-time favorite movie, a story of selflessness, sacrifice and dedication to a noble cause, but it's not preachy or boring. It just never gets old, despite my h
8	I sure would like to see a resurrection of a up dated Seahunt series with the tech they have today it would bring back the kid excitement in me.I grew up on black and v
9	This show was an amazing, fresh & innovative idea in the 70's when it first aired. The first 7 or 8 years were brilliant, but things dropped off after that. By 1990, the shov
10	Encouraged by the positive comments about this film on here I was looking forward to watching this film. Bad mistake. I've seen 950+ films and this is truly one of the v
11	If you like original gut wrenching laughter you will like this movie. If you are young or old then you will love this movie, hell even my mom liked it. Great Ca
12	Phil the Alien is one of those quirky films where the humour is based around the oddness of everything rather than actual punchlines. At first it was very oc
13	I saw this movie when I was about 12 when it came out. I recall the scariest scene was the big bird eating men dangling helplessly from parachutes right out of the air. T
14	So im not a big fan of Boll's work but then again not many are. I enjoyed his movie Postal (maybe im the only one). Boll apparently bought the rights to use Far Cry long
15	The cast played Shakespeare. Shakespeare lost. I appreciate that this is trying to bring Shakespeare to the masses, but why ruin something so go

BLOCK DIAGRAM-



RESULT- The model correctly identifies that the given review is a positive one

TESTING THE MODEL

In [56]:
```
def predict_sentence(sentence):
    sentence = [sentence]
    sequences = tokenizer.texts_to_sequences(sentence)
    padded = pad_sequences(sequences, maxlen=max_length, padding=padding_type, truncating=trunc_type)
    sentiment = model.predict(padded)
    return ("positive",float(sentiment)) if sentiment > 0.5 else ("negative", float(sentiment))
```

In [57]:
```
df.review[4]
```

Out[57]:
```
'petter mattei love time money visual stun film watch mr mattei offer us vivid portrait human relat movi seem tell us money pow
er success peopl differ situat encount variat arthur schnitzler play theme director transfer action present time new york diffe
r charact meet connect one connect one way anoth next person one seem know previous point contact stylish film sophist luxuri l
ook taken see peopl live world live habitat thing one get soul pictur differ stage loneli one inhabit big citi exact best place
human relat find sincer fulfil one discern case peopl encount act good mr mattei direct steve buscemi rosario dawson carol kane
michael imperioli adrian grenier rest talent cast make charact come aliv wish mr mattei good luck await anxious next work'
```

In [58]:
```
predict_sentence(df.review[4])
```

Out[58]:
```
('positive', 0.9858889579772949)
```

CONCLUSION-

The model approaches an accuracy of about 80-90 %, which is pretty good. Further changes can be done, in case to improve the accuracy. The model was correctly able to predict if the given data review was a negative or a positive review.