

STOCK ANALYSIS AND MOVEMENT PREDICTION USING NEWS HEADLINES

Using LSTM MODEL based on Neural Networks in Deep learning

UNDER THE GUIDANCE OF:

MR. D.NAGA TEJ

(Assistant professor)

Dept of IT, GVPCOE(A)

Visakhapatnam

TEAM MEMBERS:

1. N Surendra (16131A1275)
2. N Gnanesh (16131A1277)
3. G SaiPraneeth (16131A1293)
4. V Bharadwaz (16131A12A6)

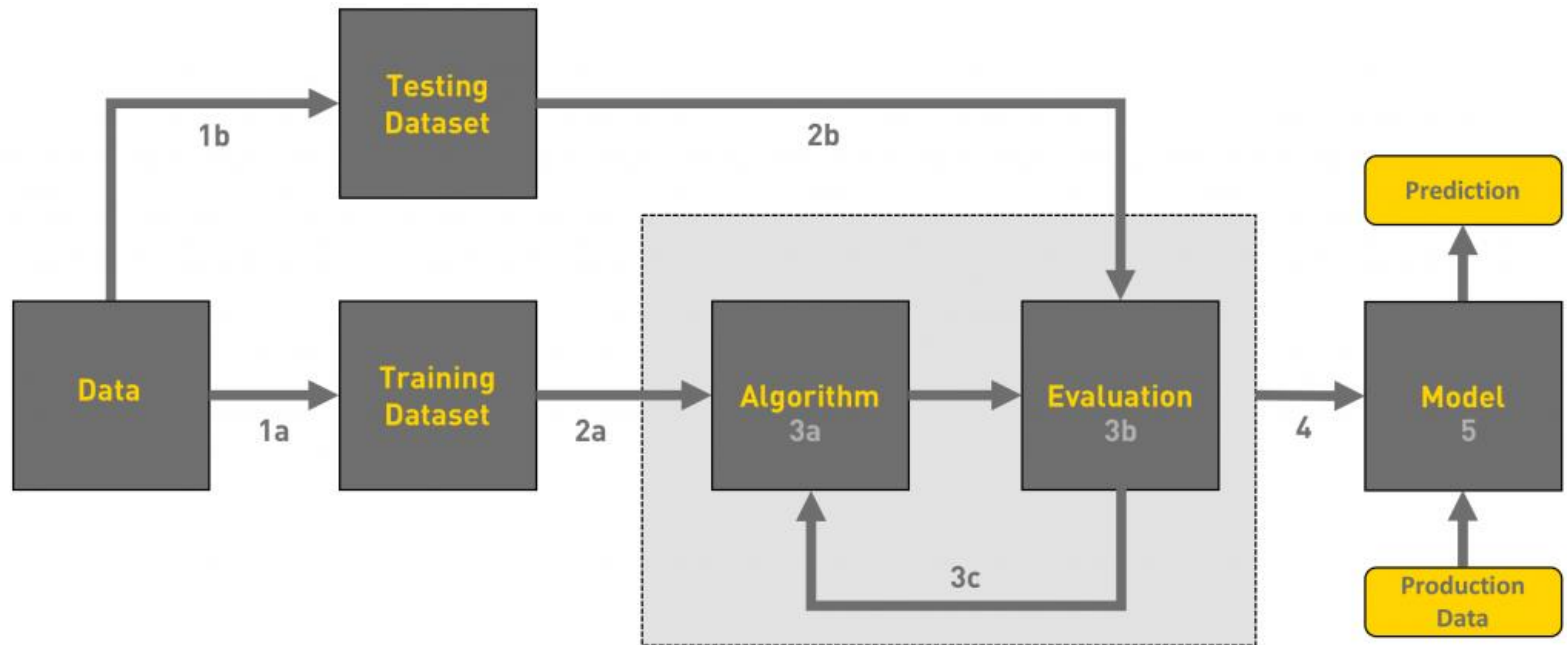
INDEX

- Abstract
- Generic ML Process Flow
- Existing Models
- Motivation
- Proposed Model
- Process Work Flow
- LSTM Architecture
- GloVe
- Word Embeddings
- Code

ABSTRACT

- Stock price prediction and its movement is extremely important for making wise investment decisions.
- Thus predicting the movement of stock prices demands an efficient and flexible model that is able to capture long range dependencies among news items and provide accurate results.
- The goal of this project is to use the top, daily news headlines from various sources to predict the movement of the Dow Jones industrial average.
- The news from a given day will be used to predict the difference in opening price between that day, and the following day.
- We would like to develop a deep learning model based on long short term memory recurrent neural network (LSTM-RNN) to train on our data and use it for prediction.

Generic ML Process Flow



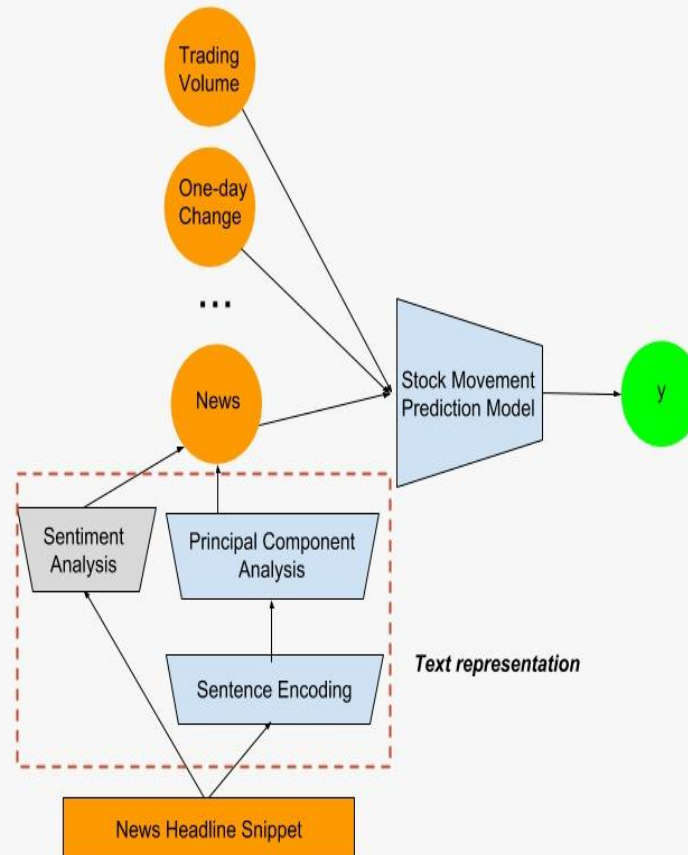
Existing Model

- There are several models which are based on prediction techniques like **Regression-Nearest Neighbours, Neural Networks** etc..
- They have their fair share of disadvantages.
- Disadvantages:
 - Regression is mostly limited relationship and easily affected by outliers.
 - Recurrent Neural Network-suffer with Vanishing Gradient.
 - The main disadvantage of the KNN algorithm is that it is a lazy learner.
 - Some models cannot depict the exact relationship between nonlinear attributes.

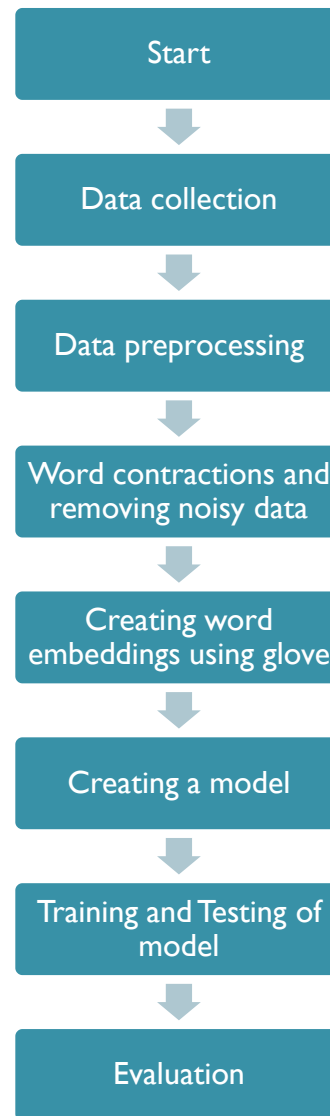
MOTIVATION

- To Predict The Stock Movement For Each And Every Day.
- To Improve The Accuracy Of Existing Mechanisms
- To Help The Users In Making Wise Investment Decisions I.E. Which Stock To Invest On, When To Invest Etc..
- We Decided To Use LONG SHOT TERM MEMORY(LSTM) So That It Overcomes All The Disadvantages And Improves Accuracy.

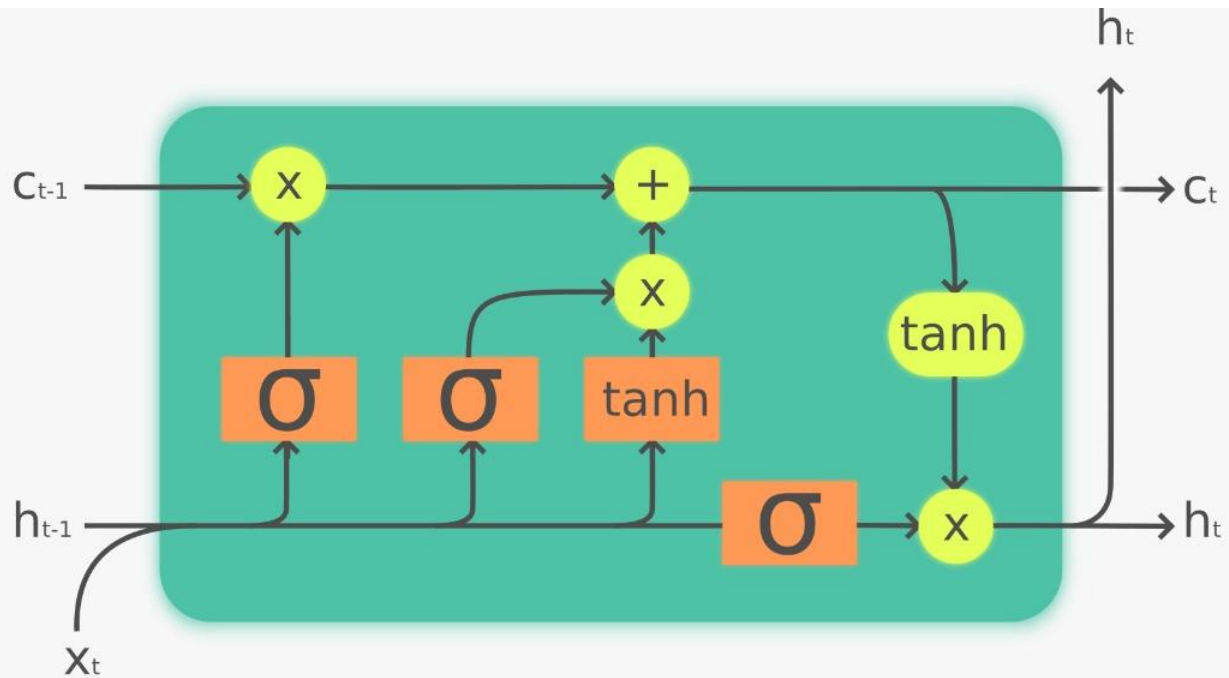
PROPOSED MODEL



Process Work Flow



LSTM ARCHITECTURE:



Legend:

Layer



Pointwise op



Copy



GloVe: Global Vectors for Word Representation

- GloVe is an unsupervised learning algorithm for obtaining vector representations for words.
- Training is performed on aggregated global word-word co-occurrence statistics from a corpus resulting in linear substructures.

Highlights:

1. Nearest neighbors
2. Linear substructures
3. Training

WORD EMBEDDINGS

- "Word embeddings" are a family of natural language processing techniques .
- It aims at mapping semantic meaning into a geometric space.
- This is done by associating a numeric vector to every word in a dictionary, such that the distance (e.g. L2 distance or more commonly cosine distance) between any two vectors would capture part of the semantic relationship between the two associated words.
- The geometric space formed by these vectors is called an *embedding space*.

Word Embeddings using GloVe

- It's a somewhat popular embedding technique based on factorizing a matrix of word co-occurrence statistics.
- we will use the 300-dimensional GloVe embeddings of large number of English words computed on a dump of English Wikipedia.
- APPROACH:
 - Convert all text samples in the dataset into sequences of word indices.
 - Prepare an "embedding matrix" which will contain at index i the embedding vector for the word of index i in our word index.
 - Load this embedding matrix into a Keras Embedding layer.
 - Build on top of it a neural network.