

# PROJECT SYNOPSIS

## MOVIE RATINGS SENTIMENT ANALYSIS

(Tentative project Name)

*Submitted towards the partial fulfillment of the criteria for award of Post Graduate  
In Data Analytics by Imarticus*

*Submitted By:*

*HARIHARASUDHAN P (IL031248)*

*Course and Batch: Post Graduate in Data Analytics & Feb-  
2022(PGA-22)*



## **Scope & Objective:**

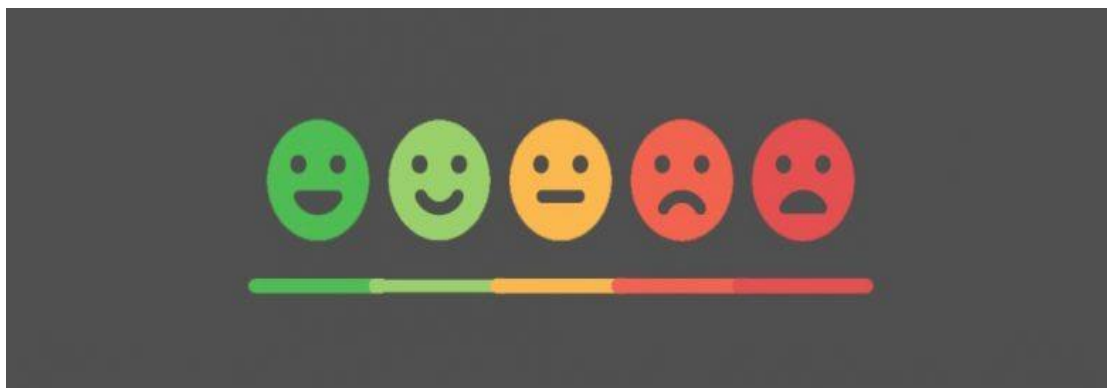
The World Wide Web such as social networks, forums, review sites and blogs generate enormous heaps of data in the form of user's views, emotions, opinions and arguments about different social events, products, brands, and politics. Sentiments of users that are expressed on the web has great influence on the readers, product vendors and politicians. The unstructured form of data from the social media is needed to be analyzed and well-structured and for this purpose, sentiment analysis has recognized significant attention. Sentiment analysis is referred as text organization that is used to classify the expressed mind-set or feelings in different manners such as negative, positive, favorable, unfavorable, thumbs up, thumbs down, etc. The challenge for sentiment analysis is lack of sufficient labeled data in the field of Natural Language Processing (NLP). And to solve this issue, the sentiment analysis and deep learning techniques have been merged because deep learning models are effective due to their automatic learning capability. This highlights latest studies regarding the implementation of deep learning models such as deep neural networks, Recurrent neural networks and many more for solving different problems of sentiment analysis such as sentiment classification, cross lingual problems, textual and visual analysis and product review analysis, etc.

## **Keywords:**

Sentiment analysis; recurrent neural network; deep neural network; convolutional neural network; recursive neural network; deep belief network.

## **Business Problem Statement:**

The problem in sentiment analysis is classifying the polarity of a given text at the document sentence, or feature/ aspect level. Whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative or neutral.



The conventional approach to sentiment classification involves several steps, from structuring the text data to understanding the customer sentiments. Over the years, Deep Learning has transformed Sentiment Analysis to a whole new level. With the introduction of Transformers and Transfer Learning, building a model for sentiment classification is a mere matter of minutes. However, knowing the basics of sentiment classification always comes in handy.

### **Data Sources:**

1. The repository that we used for dataset is Kaggle.
2. There are csv files which contains 40000 instances and 2 attributes.
3. The data set from Kaggle repository contains attributes of each state details including their Text, Label. This dataset is taken as the training data.

### **Analytics Tools:**

- **Jupyter Notebook:** collaborative work capabilities.
- **Pandas:** A python data analysis library enhancing analytics and modelling.
- **Matplotlib, Seaborn:** A python machine learning library for quality visualizations.
- **Tensor flow and Keras:** A python that enables us to estimate and analyze various models.

### **Analytics Approach:**

- To create algorithms to comparative study among the proposed technique and check the performance of the model with analytics.
- Sentiment analysis evaluation has been conducted using conventional approaches with help of movie rating. These approaches involve manual collection and assessment of raw data. The traditional approaches for sentiment analysis is a type of market analysis that includes the use of text analysis, biometrics, natural language processing (NLP), and computational linguistics to recognize the state of the said information.

## **KPIs, Timelines, Milestones(proposed):**

### **KPIs:**

#### **For Classification:**

- Accuracy
- Precision
- Recall
- F1-Score
- Cohen's Kappa

### **Timelines:**

Expected to complete the project by 15-08-2022

### **Milestone:**

Expected processing and making prediction based on the dataset achieve 90%.

### **File Format:**

The datasets are provided in CSV format, with the following features:

Retrieved from kaggle:

- 1) Text
- 2) Label