**Capstone Project 2: Sentiment Analysis**

We will visualize the movie reviews that people have put forward. And also based on their reviews we will be building a Prediction Model using different machine learning and deep learning models which signifies whether the sentiment is positive or negative for the product.

This model can be used by movie industries and streaming partners on how the sentiment of the movie is on their platforms helping them to take further decisions on how to keep customer engaged to their platforms

We will IMDB 50k Movie review dataset for this project. The dataset contains 50000 reviews and their labelled sentiment (positive or negative) .The dataset is available in Kaggle website.

We need to pre-process the text in the reviews removing un-necessary tags, punctuations and spaces.

We need to then remove the **stop words** and they form the major chunk of data and from which we will not derive any sentiment and it is generally used across all reviews.

We will applying **countVectorizer** and **td-idf Vectorizer** to convert the text data into numerical data before feeding into the model.

We will perform train test split and build on logistic regression, Multinomial Naive Bayes and linear support vector machines models and compare their performances

We will perform deep learning models like LSTM on the dataset to achieve better results.