**BSE-01 CONCEPT PAPER**

**ANALYZING STAR WARS MOVIE SCRIPTS TO COME OUT WITH PREDICTIVE MODELS.**

# INTRODUCTION

The project will be studying and analyzing the movie scripts of Star Wars from Episodes IV, V, and VI.

1. Episode IV: A New Hope – a 1977 American epic space opera film written and directed by George Lucas.
2. Episode V: The Empire Strikes Back – a 1980 American epic space opera film directed by Irvin Kershner.
3. Episode VI: The Return of the Jedi – a 1983 American epic space opera film directed by Richard Marquand.

At the end of the project, we will be able to come out with character’s opinions, sentiments, and emotions as well some clear understanding on the movie itself.

# BACKGROUND

Star Wars is a popular film franchise that takes place in a galaxy far, far away. This is a collection of script dialogue between characters for the first three movies (episodes 4-6). We are going to perform text mining on the script dialogue so as to get some clear and clean data about the movie. The resulting data will then be subjected to some analytic reviews, probably using Machine Learning technique. We will then be able to come out with some predictive works and visualization.

# GOAL

The main goal of the project is to mine all the text in the three collections of the characters dialogue, and perform sentiment or opinion analysis and the data generated. This will be achieved using one of these techniques, machine learning or Naïve Bayes.

# OBJECTIVES

* Explain the purpose of the project
* Perform text mining
* Run sentimental analysis using ML
* Predict the characters attitude and what the movie is pointing to.

# METHODOLOGY

The project will be going through three major steps, namely; test mining, sentiment analysis, evaluation and conclusion.

* Text mining

We will be transforming all the text into data for easy and effective analysis. We will need to as well do data cleaning to make sure all our data is accurate.

* Sentiment analysis

This step will aim at depicting and classifying how positive, or negative each character is as well as the neutrality of the character. We will also be able to determine the intended opinion of all characters.

* Evaluation and conclusion

The third and the last step of the project will be evaluating all the results of our analysis. We will then be able to sport the character who is so much positive, neutral or negative. We shall, in each step, come out with some graphical visualisations to get some insites on all the data that we shall have. We shall then make our conclusion basing on the results that we will come out with after analysis.

## OUTCOMES

We expect to design some meaningful and clear results after the project, basing on the problem we have. Also, after the project, we are so optimistic that we will be smart, cute and perfect data analyst/scientist.

The project will be complete by the end of JULY, 2018 when we will be making the final presentation to our cute and lovely instructor Mr. NOAH KANGE. At that time, we will decide whether to sit back and celebrate our success or continue with other projects that will be at hand.