

# QUALITATIVE ANALYSIS AND TOPIC MODELLING ON TED TALKS USING NLP

## INTRODUCTION

In this project, I have tried to do topic modelling on a dataset containing TED Talks ([<https://www.kaggle.com/datasets/jeniagerasimov/ted-talks-info-dataset?rvi=1>]) using LDA (Latent Dirichlet Allocation). I have also tried to judge the quality of each topic using the number of likes and views talks on each topic have.

## TECH USED

I have used the following python libraries: -

1. Pandas for data handling
2. Spacy for tokenisation and lemmatisation of text
3. Gensim for the LDA model
4. matplotlib for plotting data
5. Wordcloud for topic visualisation

## INSTRUCTIONS TO RUN

Follow along the steps in the Jupyter notebook – main.ipynb.

THE OBSERVATIONS AND CONCLUSIONS OF THE ANALYSIS ARE ALSO PRESENT IN THE JUPYTER NOTEBOOK