Springboard Capstone 1 : TED Talk Text Analysis

Pavan Poosarla; [pavanpoosarla01@gmail.com](mailto:pavanpoosarla01@gmail.com) ; Sep, 2019-Feb, 2020

# Introduction

The goal of this capstone project is to automatically assign tags to a TED talk based on its transcript. To do this, we train our model on the existing TED talk dataset. TED talk dataset is taken from the Kaggle repositories.

Kaggle TED talk transcripts : <https://www.kaggle.com/rounakbanik/ted-talks/version/3>

This data has been scraped from the TED talk official website and is available now by a creative commons license. The database consists of talks scraped till Sep 21st, 2017. It has two files,

**ted\_main.csv :** Metadata associated with each

**transcripts.csv :** It consists of a file consisting of the text of the transcript along with the corresponding url of the talk. Few comments on the transcripts

# 