

Springboard Data Science Capstone Project - Predicting User preferred news articles

Ramaa Vissa

Mar 22, 2020

Contents

Introduction

Problem: “what is the favorite news article that a particular user will read”

Data

Data Cleaning

Split the data into train and test data

Model

1. Introduction

Most manufacturing factories have a quality control unit that oversees the predictive maintenance. On the floor, machine failures can occur due to wear and tear of the machine over prolonged use. The problem is to capture the failed unit before it can cause more failures downstream.

The data presented here is obtained from simulation methods and this is a use case for machine predictive maintenance.

We would like to implement a machine learning algorithm to detect a bad device during a particular time frame and prevent problems before they occur.

This is a Multi-class classification problem. A machine learning algorithm is used to create the predictive model that is trained from historical data.

Problem: “what is the favorite news article that a particular user will read”

- label/predictions is the news articles - video id.
- predictors/independent variables:

These are the text topics and text articles,
category_name,story_text,short_description.

2. Data

The data has text columns, video_id,
story_text,category_name,short_description.

The acquired data sets were created by the actual newspaper company.

3. Data Cleaning

There were no missing values when I inspected the data.

4. Split the data into train and test data

In this Use case i assumed that the user would be interested in a particular video id. Using that I trained the model using cosine similarity.

5. Model

This is a Text classification problem. TF idf was used to convert words to vectors, cosine similarity was used for model prediction.