ISBESTSELLER PREDICTION

PREPEARD BY:
RANYA ALKHTANI

EMAIL & GITHUB ACCOUNTS:
RANYAALKHTANI@GMAIL.COM

CLASSIFICATION PROBLEM
PROJECT REPORT

ABSTRACT

The goal of this project was to make classification on sentiment analyzed data in order to predict whether the course will be bestseller or not. I worked with Udemy courses Dataset which I scribed. after exploring the data I come up with a target and an idea of model. I wanted to make prove of concept using Udemy courses Dataset. the classifier that I chosen as the Best model, have been going through a lot of experiences with different features and algorithms to improve its performance.

DESIGN

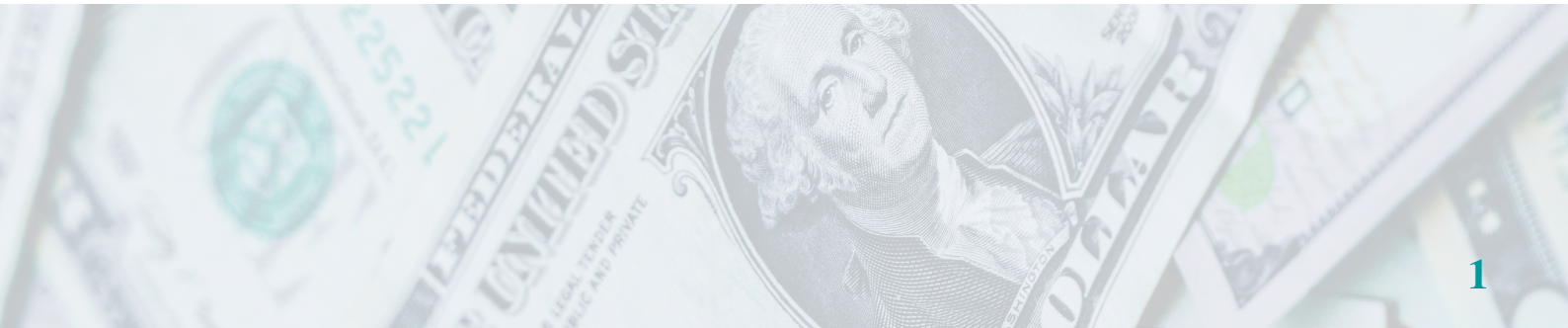
This model is to provide the courses required in the market place, especially as they rely on the specific criteria to meet and succeed in this task. Predict whether the course is a best seller or not may be one of the most important criteria that motivates an instructors to offer that course. By using AI-based classification models, we can help the teacher to cover such a need.

DATA

The data obtained by performing web scarbing on (https://www.udemy.com) website which include many courses with 65 languages and covering almost all subjects that anyone needs. The dataset consists of approximately 20,000 entries linked to 11 features. The features we concern with and get from this website for the courses are ["Title", "Instructor", "Description", "Rating", "Participants", "Original_Praice", "Discount_Price", "Hours", "num_lectures", "Level", "Catogory"].

ALGORITHMS

- **Data Collection:** web scribing.
- **EDA: plotting :** null, correlation.
- **Data Cleaning:** Handling null, stop-word removal, punctuation removal, text lowercase, numbers removal.
- **Pre-Processing:**
  - **Feature preparation:** categorical feature encoding.
  - **Feature Engineering:** Tf_idf feature, word2vec.
- **Modeling:** Logistic Regression: KNN, Naive Bayes, Gradient Boosting, Random Forest.



TOOLS

Technologies: python, google colab
Libraries: SKlearn, StatesModel,neighbors, svm, BeautifulSoup, Selenium, NumPy, Pandas, Matplotlib, Seaborn, langdetect, LinearRegression, time, os, Request, IPython.core.display.

COMMUNICATION

in this part of the report I will show and document all of communication work has been shown both of visualisations and slide show: