Related Works

[4] Play Store

**Methodology for Analyzing the Traditional Algorithms Performance of User Reviews Using Machine Learning Techniques**

As mentioned earlier, google play is one of the most used android applications in the world. The [4] is an in-depth research study on the reviews scrapped from the app reviews of the google play store. The authors implemented ML models to classify the semantics of the text as negative, positive, or neutral. The study was able to verify Logistic regression as the better model based on F1 score and accuracy. The limitation of the paper is one dimension focus on text reviews and feedback while not considering the fact that most users prefer to leave start rating. The Google Play Store gives the point-based feedback system. The comment is often given by users with negative experience. This makes study biased.

[5] Play Store

# Automatic movie ratings prediction using machine learning

While the previous paper focused on text reading, [5] is research paper with direct study topic with paper. The study is conducted to predict movie user ratings using different ML methods. The study showcased that Latent Variable Model (LVM) is the better model for predicting audience ratings for movies. While the paper is successful on implementing ML methods that can predict ratings, the usual database used is limited to one region and lacks diversity to show the large-scale implementation.

[6] play store

# An Efficient Deep Learning Model to Infer User Demographic Information From Ratings

The [6] discusses one of the benefits of user ratings which is inferring the demography of the audience of the user. This is a useful feature to provide directed services and products. The paper discusses the potential new model for this specific type of ML prediction.

[7] play store

# Predicting numeric ratings for Google apps using text features and ensemble learning

The [7] elucidates about the discrepancy between numeric rating and reviews. The paper applies ML model on the average semantics to portray the numeric ratings and then compare them with the source numeric ratings. The classification model used here is random forest classifier. The study shows that User Ratings are often manipulated and do not coincide the true user experience. This is a useful topic for further research since the project aims to predict the user rating on Google apps. It is thus necessary to make sure that the rating are genuine and not manipulated by anyone.

[8] play store

# A study of methods for normalizing user ratings in collaborative filtering

The [8] focuses on the user behavioral pattern for the user ratings. The study suggests that some users are more inclined to give ratings as compared to others and hence it is important to normalize for unbiased study. The paper uses Gaussian Normalization and Collaborative Filtering method to reach conclusion on the research question. The paper concludes with Collaborative Filtering more accurate as compared to other.

[9] Sunroof

# Predicting solar generation from weather forecasts using machine learning

Solar energy has had deep research for the past few decades. The amount of related works on solar energy is plentiful. As such,[9] discusses the prediction solar power generation from weather forecast. This is in-line with our study that also use the average sunlight received as a predictor variable.

[10]solar

# Predicting Solar Radiation Using Machine Learning Techniques

# The [10] is a research study that used various ML models and compare the results in order to find the best predicting model to Solar Radiation. The study concludes with SVM ass the better model for the research.

# [11] solar

# [11] Machine learning models to quantify and map daily global solar radiation and photovoltaic power

While most studies prefer traditional models for solar energy prediction, [11] uses Extreme Learning Machine (ELM) to map daily global solar radiation. The paper studies photovoltaic power quantification. The lack of ELM methods in python proved to be an hinderance in this project. The inability to compare random forest model with ELM needs to be done at a later stage.

[12] solar

# Prediction of solar energy guided by pearson correlation using machine learning

Machine learning related works in Solar predictions often compare only among ML models. [12] includes Artificial Neural Networks (ANN) for comparative study with ML models namely SVM. This can be done be future evaluation of this project. The study shows the outperformance of ANN with respect to ML models.

[13] solar

# A city-scale roof shape classification using machine learning for solar energy applications

The project uses roof type, number of panels on roof for predicting the potential solar energy generation. This is also discussed in [13], where the classification ML model is used for classifying roof shapes in Geneva for solar energy applications. The studies agree with flat roof being the most efficient for solar energy applications as compared to other roof types.

[14]ytube

# Hybrid machine learning approach for popularity prediction of newly released contents of online video streaming services

The paper[14], I the basis of choosing XGBoost as a ML model for project. Predicting the popularity is essential for any revenue generating company which deals with content manufacturing. This also shows the most popular video category.

[15] ytube

# Online video channel management: An integrative decision support system framework

From a content creator perspective, learning the key to audience taste the most profitable aspect that they can predict. [15] discusses the decision support system that can be employed for online video channel. The study helps in topic consideration for the project.

[16] ytube

# Emotion classification of YouTube videos

An important aspect of machine learning includes the prediction of human emotions. The paper [16] classifies the videos on emotion factors. The study focuses on the ability of machine learning to predict complex prediction variables. This paper was useful for topic consideration for the project.

[17] Solar

# Improved machine-learning mapping of local climate zones in metropolitan areas using composite earth observation data in Google Earth Engine

The Project Sunroof relays great importance on the shadows cast my land features on roofs where solar panels can be constructed. [17] improves upon the machine learning aspects of Google Earth data. The project relies on the dataset from Google Earth, hence a study inspection to get a view of ML models was necessary.

[18] solar

# A machine learning approach to modelling solar irradiation of urban and terrain 3D models

City infrastructure shape sunlight for low laying houses and lands. This has implication for any solar panels that are to be constructed there. [18] investigates a ML model for urban to model solar irradiation.

[19] ML

# Comparing performance of multiple classifiers for regression and classification machine learning problems using structured datasets

The study [19] has been a significance influence on the project for choosing XGBoost as a classifier for YouTube Dataset. The study shows the superior performance of the XGBoost as compared to other models. This is inline with the project results.

[20]

# Random forest algorithms for the classification of tree-based ensemble

Random forest decision tree has the highest performance parameters in the project study. This is in-line with [20] conclusion. This is critical for the model selection for the project. The study works upon the efforts for improving the current model. There is a comparison with logistical regression.

[21]

## [**The role of demographics in online learning; A decision tree based approach**](https://www.sciencedirect.com/science/article/pii/S0360131519300818)

The project focuses much on the data generated because of the online activity. [21] elucidates on the role of demographics in online learning. The paper shares light on the decision tree machine learning model. The focus is much on the effect the different demography shed on the online activity.

[22] youtube

Can you predict the rating of an app?

<https://www.kaggle.com/almatyeraly/can-you-predict-the-rating-of-an-app>

The [22] on Kaggle discusses the same topic as the project. The work here is limited to linear regression. The author concludes that the there is not enough correlation between predictors variable to apply linear regression for a successful model. The project wants to further this study and apply other models to see if the previous works can be confirmed and added upon.