KSA Projects' Budgets Predictions

Team Techy Fighters

Saudi Arabia's vision 2030 has many huge projects that affect the region and its economy. For most of these projects, the budget is exponentially high, and as such have burdened the Kingdom's budget towards projects of different scales.

Among the objectives of the vision of the Kingdom of Saudi Arabia is to increase the percentage of domestic production and non-oil exports, enabling the startups to provide a more beneficial life for citizens that rely on other sources of income rather than oil.

To get these budgets, investors are considered one of the most important factors in building these projects. Investors need to know the budget for a project based on previous data to grab an idea of what and how they plan on implementing these ideas. Based on the investor's choices the economy is drastically changed due to the variance in budget and the performance of these individuals.

These types of differences cause wastage or shortage of money and manpower, thus affecting the economy and the plans of the project. To decrease these effects on the planning and project management, we decided to implement a thorough investigation and determine the prediction of the budgets for future projects so that everything is already known before the planning process which gives leeway and flexibility to handle new ideas and their variations.

Dataset

Saudi projects dataset imported from Kaggle website provides students good understanding to give insight to those who are concerned with effective, long-term investments, endeavors to be a leading investor while sparking an impact felt beyond Saudi borders. Studying Saudi projects dataset would maintain a leading and impactful investment strategy in line with Vision 2030, to drive Saudi Arabia's economic transformation through active, long-term investments and high standards of governance and transparency.

The plots

The interactive plots we have created are an informative histogram, stacked bar, timelines, bar chart, and scatter. These plots gave insight into how the budget has been allocated in different sectors especially in terms of the governmental sector.

While the government sector has the highest sum of budgets because they are responsible for doing giga -projects, transport, mobility schemes, and so many infrastructure developments, the stacked bar plot provides insight into how long the duration of government projects has been. It's easy to notice a govern mental sector has the highest duration project.

Approach

Once the prediction with the accuracy of the project budget is done for any kind of project it develops a portfolio of high-quality domestic, regional, and international investments diversified across sectors, geographies, and asset classes. Joining forces with top-tier global strategic partners and renowned investment managers, the great prediction study would act as the Kingdom's main investment arm to deliver a strategy focused on achieving attractive financial returns and long-term value for Saudi Arabia.

It has been applied two kinds of machine learning models approach which are Regression and Classification. For Regression models, analyzing Saudi projects to predict the budget of recent projects and determine whether the project budget is appropriate or not using machine learning regression models such as Linear Regression, Decision Tree Regressor, Random Forest Regressor, SVR, and XGB regressor. Then it has been applied the cost function for all the models to calculate how much wrong the model was in its prediction. On the other hand, Classification models were developed to help in the range the budget into three different categories high, medium, low. Utilizing Decision Tree Classifier, Random Forest Classifier, SVC, and XGB Classifier.

In the context of Machine Learning, the split of our modeling dataset into training and testing samples is probably one of the earliest pre-processing steps. The creation of different samples for training and testing helps us evaluate model performance, but the team decided to go with one type of split. It was initially created the training set by taking a sample with a fraction of 0.80 and the testing set of 0.20 from the overall rows in the pandas DataFrame.

Results

All results of regression and calcification models have been shown different results with different traini ng and testing accuracy scores. But the remarkable result has been issued was with XGB Classifier model s with training set accuracy of 96.0 % and testing set accuracy of 96.0%. The features were selected are se ctors, sector_budgets, type_project, project_area, region_project, and status_project. The project area has shown the most important feature. According to cost function records, the best iteration predictions of me an budget absolute error was 212824308.35 and mean budget square absolute error was 4.232462510583 03e+17 above or below the actual price since the mean budget for a project in our data is 7500842000.

Conclusion and Further Work

Filling the null values with the actual real number of the project budget column instead of the mean, median, or mode will lead the model prediction to be more accurate. However, looking for other resource s of Saudi projects and finding new important features shell increase the accuracy of the prediction.