TECHY FIGHTERS

KSA PROJECTS' BUDGETS PREDICTIONS

Machine Learning Capstone Project to predict projects' budgets

Overview

Dataset and Dashboard

Regression Model

Classification Model

Web Application

BUSINESSE PROBLEM

IN SAUDI ARABIA

BUDGET

The budgets of large projects have always burdened the Kingdom's budget

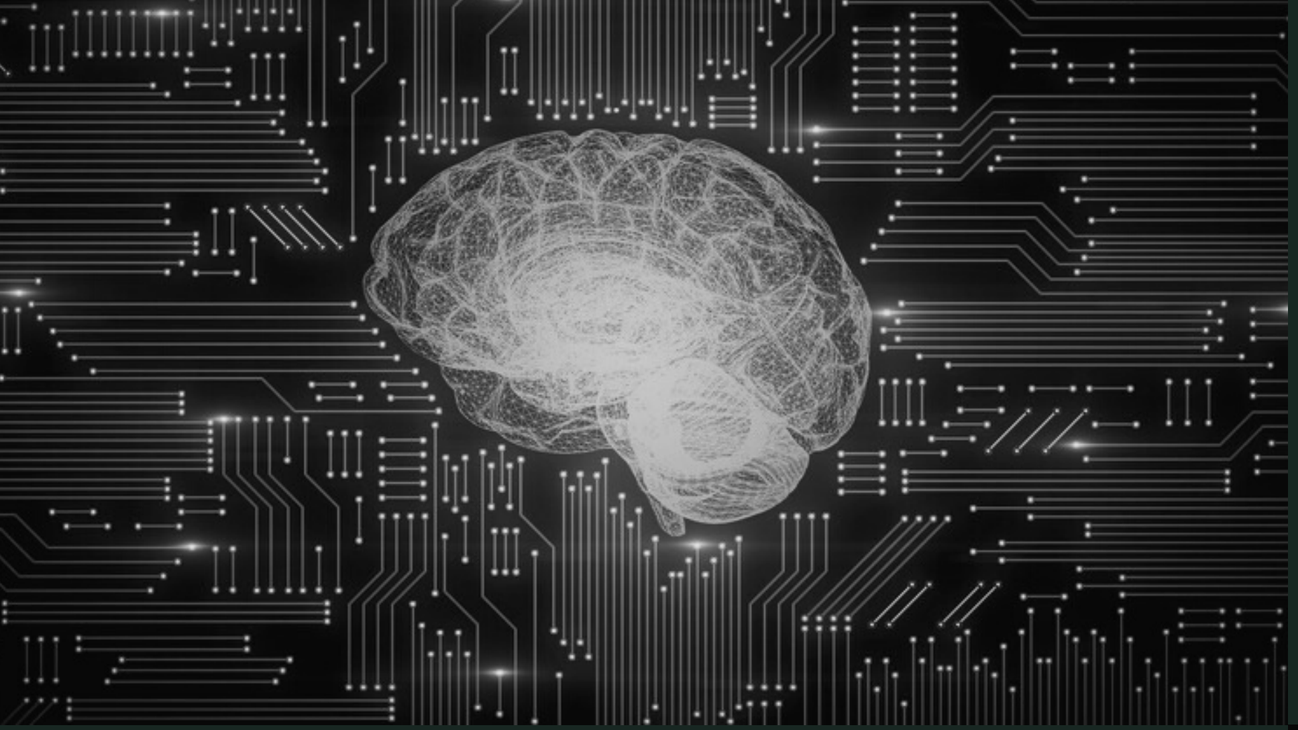
INVESTORS

Investors need to know the budget for a project based on previously existing data

ECONOMY

this causes wastage or shortage of money and manpower, and thus the economy is affected

OUR APPROCH



REGRESSION

Analyzing Saudi projects to predict the budget of recent projects and determine whether the project budget is appropriate or not using artificial intelligence and machine learning.

CLASSIFICATION

Classification to help in range the budget as three different categories high, medium, low.

Saudi Projects Dataset

CHALLENGES

THE DATASET

TRANSLATING

The only source for our data was mainly in Arabic.
To make the data usable we started by translating the columns.

CLEANING

The Area column was combination of the numeric value and the unit of measurement which wasn't unified.
we seperated the numbers from the unit and converted all area to square meters.

MISSING VALUES

The biggest challenge was the huge amont of missing values and having only one source for our data.
We used only available data in budget and imputated the rest with two methods:
the mean and KNN imputer.

Dataset Features

SECTOR

Our data contain the sectors for each project and the sector budget and the project type

DATE

We have the start and end date for each project. We extracted the month, year and duration from these columns.

AREA & REGION

The area in square meters and there are different regions.

STATUS

The projects are in different status, some are completed and some is just announced.

KSA PROJECTS' BUDGETS PREDICTIONS

Machine Learning Capstone Project to predict projects' budgets

Introduction

Saudi projects have increased recently to contribute to the relaization of vision 2030. For most of these projects the budget is exponentially high and they have always burdened the Kingdom's budget.

Objective

Build a machine learning model to predict the budget of large saudi projects.

BUSINESSE PROBLEM

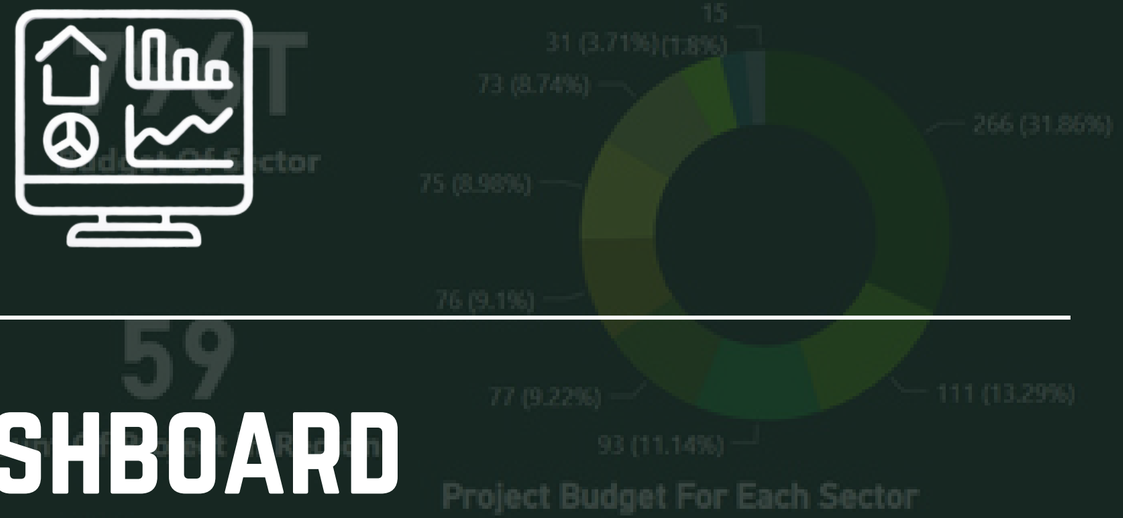
Aspecting three factors :

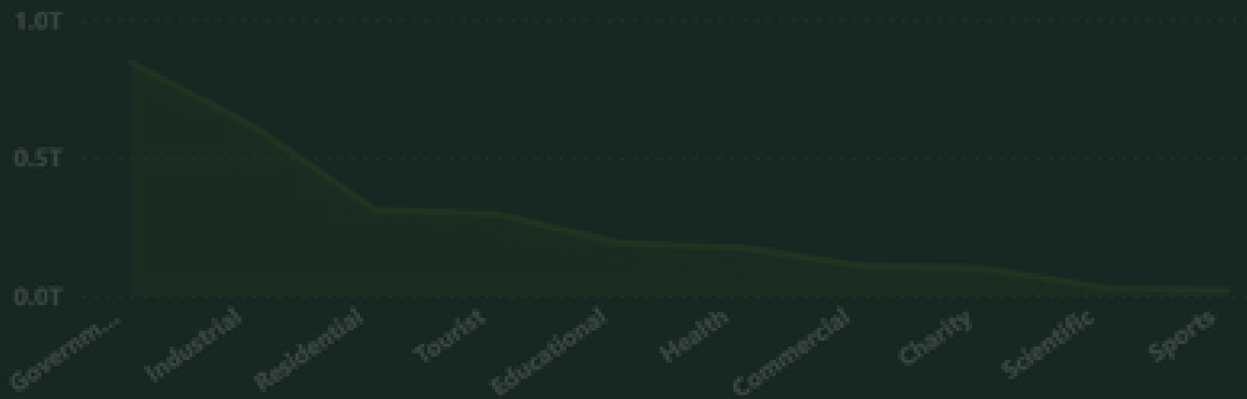
Budget

Investors

Economy

Visulization





Solution

Regression

Classification

DASHBOARD

The Models

REGRESSION

- Linear Regression
- Decision Tree regression
- Random Forest regression
- Support Vector regression
- ✓ XGB regression

[33]:
```
# Cost Function for all the models
model_name = ['linear', 'Decision Tree', 'Random Forest', 'Support Vector', 'XGB regressor']
model_pred = [preds_lin, preds_tree, preds_forest, preds_svr,preds_xgb]

for x in range(len(model_pred)):

    reg_cost(model_name[x],y_test, model_pred[x] )
```

```
Cost functions for the linear regression is:
Mean Square Error: 7.757141647892858e+19
Mean Absolute Error: 4385687863.46


Cost functions for the Decision Tree regression is:
Mean Square Error: 4.435618145589347e+19
Mean Absolute Error: 2921435460.36


Cost functions for the Random Forest regression is:
Mean Square Error: 7.131113902566983e+18
Mean Absolute Error: 991374043.95


Cost functions for the Support Vector regression is:
Mean Square Error: 8.776748878653682e+19
Mean Absolute Error: 3085830004.05


Cost functions for the XGB regressor regression is:
Mean Square Error: 4.23246251058303e+17
Mean Absolute Error: 212824308.35
```

Mean budget absolute error

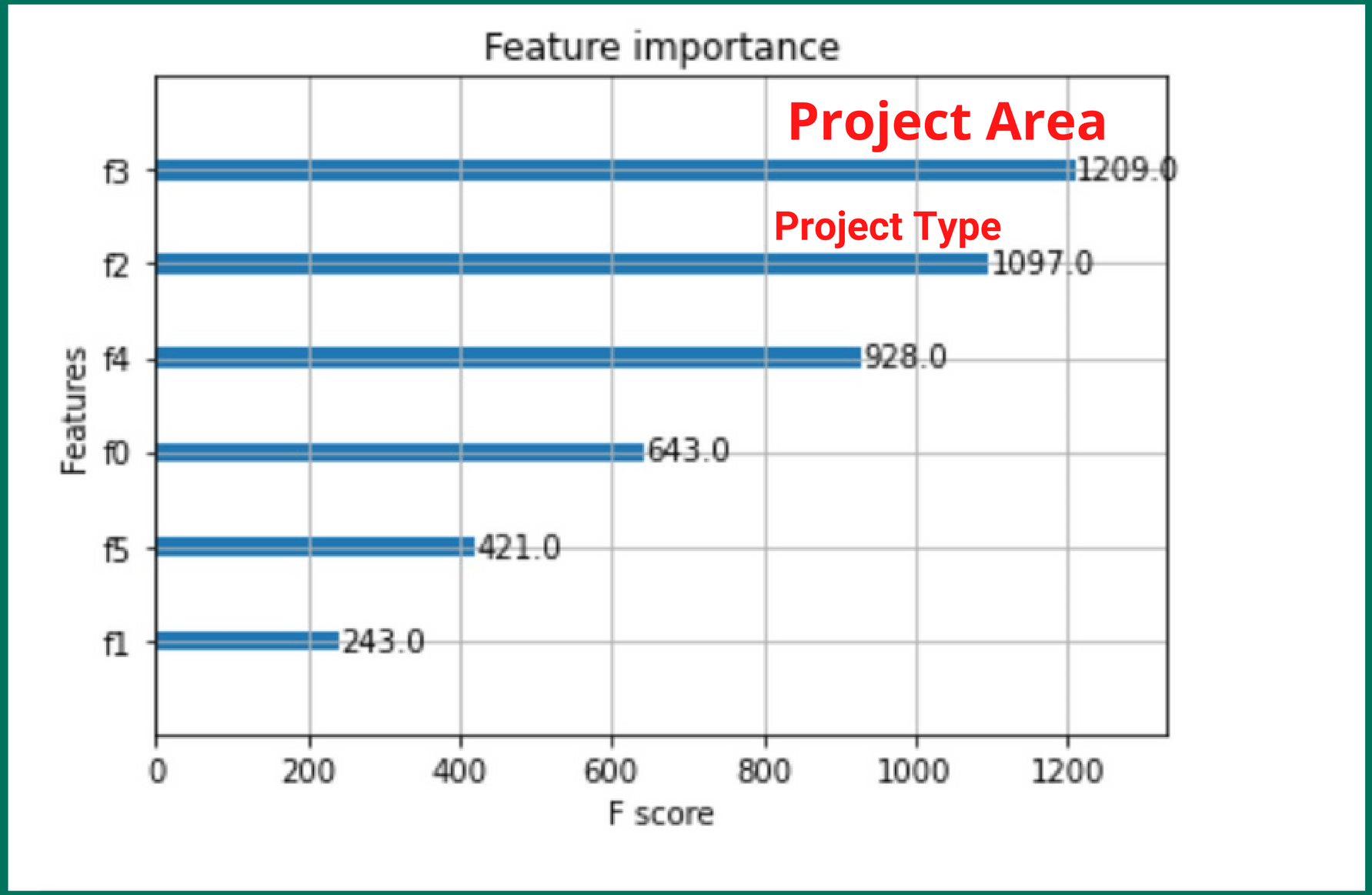
212,824,308.35

The mean budget for a project

7,500,842,000

Prediction < 20%

Accepted !


Feature importance
Project Area
Project Type
f3
f2
f4
f0
f5
f1
1209.0
1097.0
928.0
643.0
421.0
243.0
Features
0
200
400
600
800
1000
1200
F score

CLASSIFICATION

- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Classifier
- ✓ XGB Classifier

IMPLEMENT CLASSIFICATION MODELS

NEW COLUMN

defining budget column for classification, and then, we create a mask to select rows where the condition is met.

CREATES A DATAFRAME

With new column budget_project_2 which is an array of 3 values, low, medium, and high.

SPLIT DATASET

Split Dataset to train and test.

ORDINAL ENCODER

Transform the values of the column "sectors" in x_train and x_test into an ordinal value.

STANDARD SCALER

fits and transforms the data using the fit_transform method of the StandardScaler class.

MODELS

```
Accuracy: 56.0000000000
Accuracy: 56.0000000000
```

DECISION TREE CLASSIFIER

```
Training Accuracy: 94.0%
Test Set Accuracy: 94.0%
```

RANDOM FOREST CLASSIFIER

MODELS

```
Training Accuracy: 50.0%
Test Set Accuracy: 50.0%
```

SVC

```
Training Accuracy: 96.0%
Test Set Accuracy: 96.0%
```

XGB CLASSIFIER

Web App

SAVE THE MODEL

project_app.py

```
1 import pickle
```

USE STREAMLIT

```
5 import streamlit as st
```

TAKE USER INPUT

```
51 # Take the users input
52 sector = st.selectbox("Select a Sector", df_s['sectors'].unique())
53 region = st.selectbox("Select the Region", df_r['region'].unique())
54 status = st.selectbox("What is the project status?", df_st['status'].unique())
55 area = st.slider("What is the project Area in squered kilometers?", 1, 13750)
```

CONVERT AND SAVE

```
# store the inputs
features = [p_sector, p_sbuget, p_type, p_area , p_region, p_status]
```

PREDEICT USING MODEL

```
# Get prediction from the models
if st.button('Predict'):                    # when the submit button is pressed
    # Reggresion model
    prediction =  loaded_model.predict(final_features)
    st.success(f'The project budget would be: ${round(prediction[0],2)}')
    # Classification model
    prediction_c = loaded_model_c.predict(final_features)
    d = {0 : 'low', 1 : 'medium', 2 : 'high'}
    preds = d.get(prediction_c[0])
    st.success(f'The project budget would be: {preds}')
```

KSA Projects' Budgets Predictions

Calculating the budget based on input attributes

Select a Sector

Charity

Select the Region

ABHA

What is the project status?

Announced

What is the project Area in squered kilometers?

1

1 13750

Predict

What can be Improved

CONSTRUCTION COSTS

the manpower being put in these projects.

PRICES OF MATERIALS

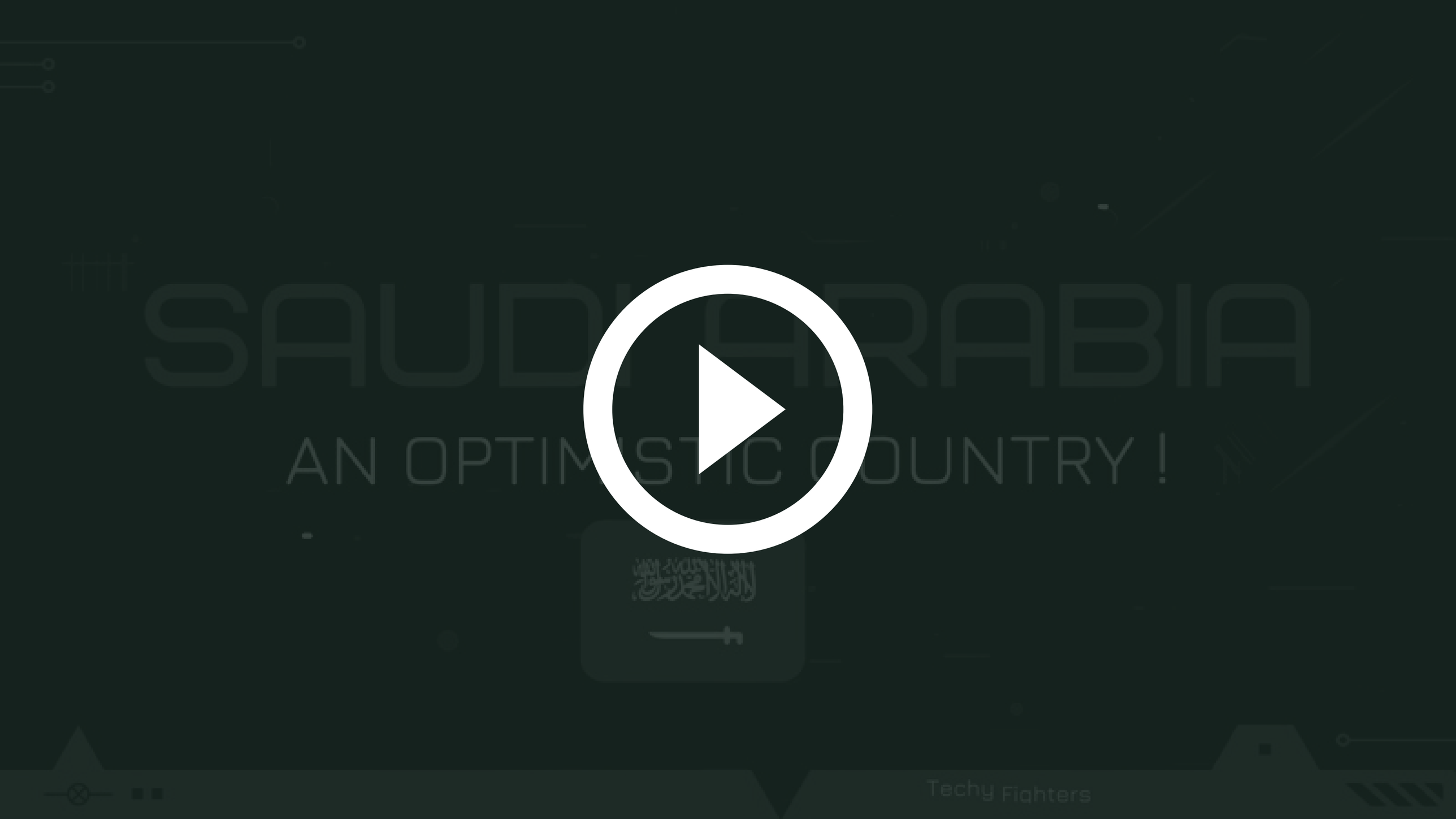
prices of the materials being used from equipment to the basic building blocks.

LEADERS OF THE PROJECTS(COMPANY BASED)

What companies are responsible for the project since each company has its own methods of implementing and designing each individual project.

OUTSOURCED OR NOT (LABOR)

labor was outsourced or not since it greatly affects the budget of the project.


SAUDI ARABIA
AN OPTIMISTIC COUNTRY !
Techy Fighters

THANK YOU FOR LISTENING

الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

Techy Fighters

CODING DOJO