iAQAR

Your confidence prediction!

T5 DATA SCIENCE BOOTCAMP

By:

Naif Albader

Yazeed Musallam
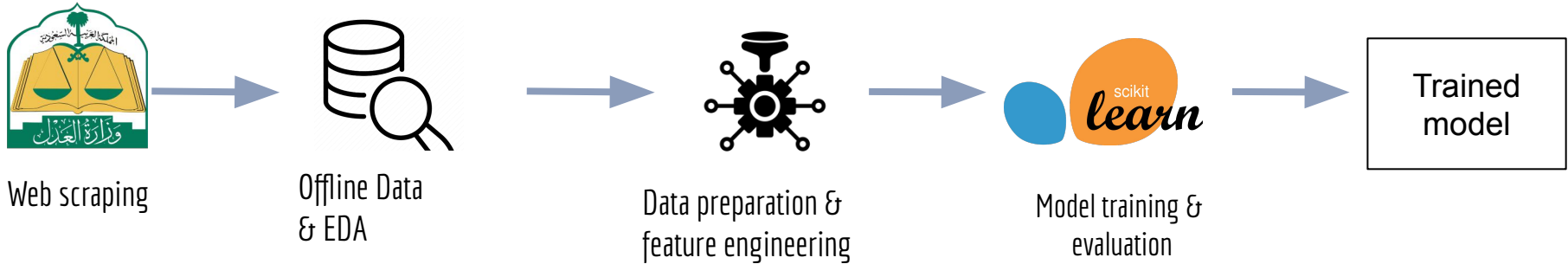
Introduction

- **What is iAQAR?**

- **Problem Statement**

- **Value of iAQAR company:**
    - What is the housing price in riyadh in the next 5 years?
    - Is the price increasing each year?
    - What is the price growth for each year?

iAQAR

iAQAR Workflow

Web scraping → Offline Data & EDA → Data preparation & feature engineering → Model training & evaluation → Trained model

Example of sale

المجموعات العقارية ا...	تفاصيل الصفقات ا...	تفاصيل الصفقات المخطط	تفاصيل الصفقات عدد ا...	تفاصيل الصفقات التار...	تفاصيل الصفقات (رقم...	السعر بالريال السعودي	المساحة بالمتر المربع	سعر المتر المربع
أحد	مخطط/2485	قطعة 257/1	2019-01-02	7937108	ر.س. 300,000	410.40	731	
أحد	مخطط/2485	قطعة 259/1	2019-01-21	8007807	ر.س. 300,000	409.80	732	

Example of a sale from ministry of justice website
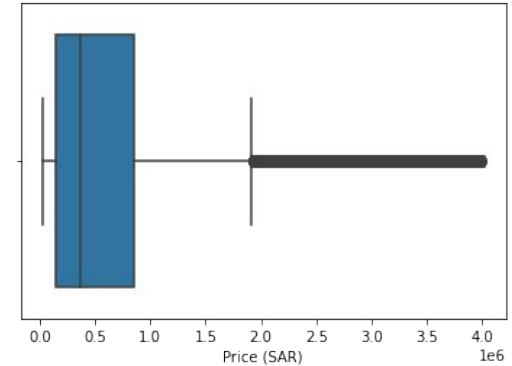
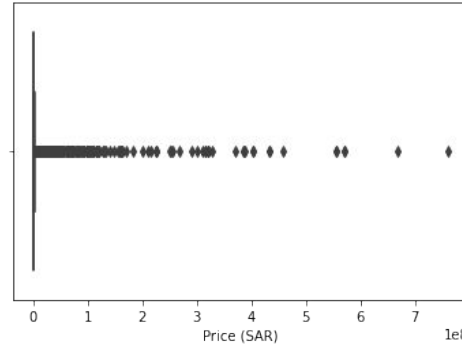iAQAR 4

Web Scraping & EDA

Data scrapped:

- **Years:** 2010 to 2021
- **Number of deals:** 600K
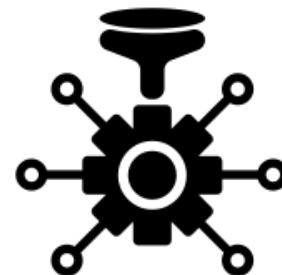- **Number of features:** 8

EDA:

- **Outliers are handled (IQR)**
- **1883 Duplicated deals**
- **No nulls are found**

Data preparation & feature engineering

Features add:
- Change date to numeric value
- Year
- Yearly/Area
- Categorical values are changed to one label encoder

	Date	Price (SAR)	Area (m2)	Price_log (SAR)	year	Yearly_Area	N0	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
3	733781	150000.00	500.00	11.92	2010	4.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	733774	662860.00	1441.00	13.40	2010	1.39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	733781	130000.00	625.00	11.78	2010	3.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	733798	170000.00	750.00	12.04	2010	2.68	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	733796	140000.00	625.00	11.85	2010	3.22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5 rows × 189 columns

iAQAR

Model Training & Evaluation

Data split:
- **Training:** 2010 - 2020
- **Validation:** 2021/1 to 2021/5
- **Testing:** 2021/5 to 2021/9

Results on test set (Neural Network)
- **R2 (Training):** 0.70
- **R2 (Validation):** 0.62
- **R2 (Testing):** 0.60

Model training & Evaluation

Tested models:

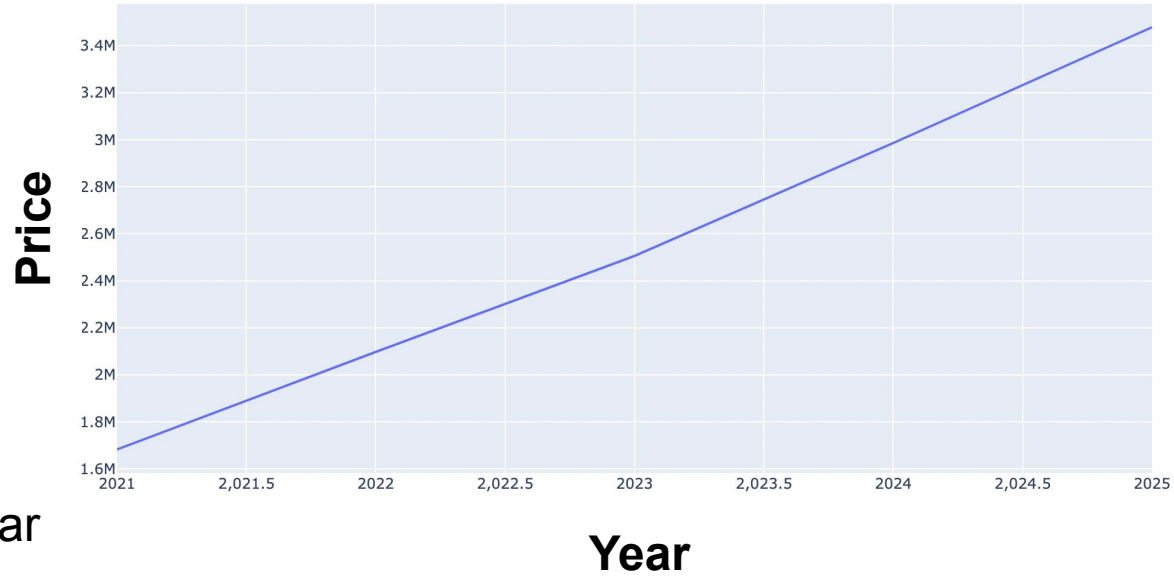| Model	R^2
Linear regression (Baseline)	0.47
Polynomial regression	-0.22
Random forest	0.58
XGboost	0.54
Neural network (The best)!	0.60

iAQAR

Sample test

Sample Data:
- **Area:** 500
- **Date:** 2021-09-16
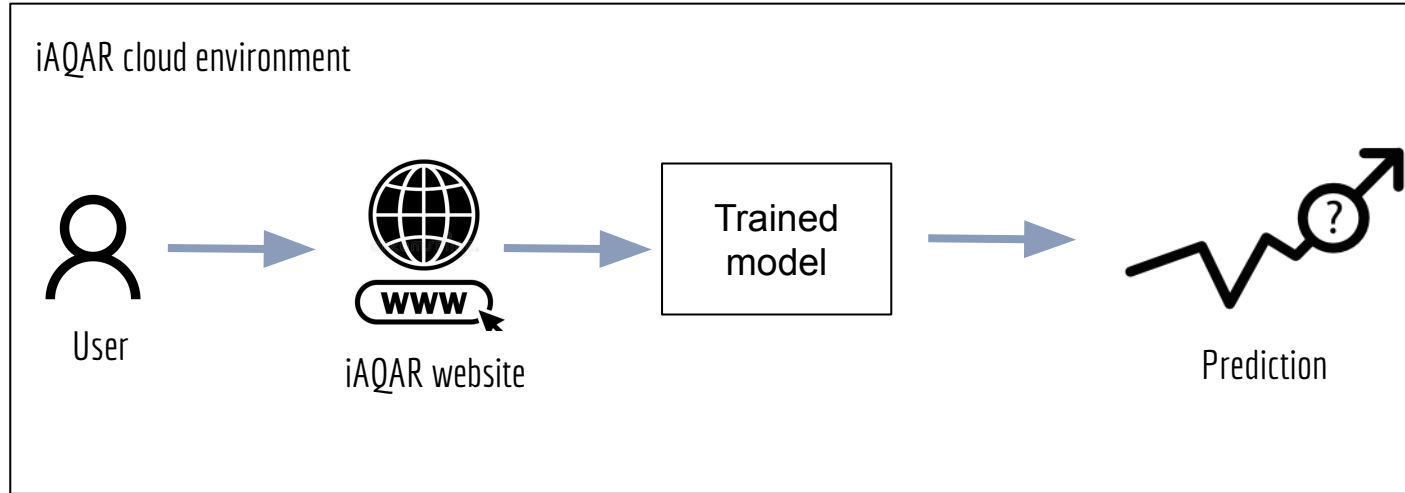- **Neighbourhood:** "الياسمين"

Observation:
The price increases every year

Sample of land

iAQAR

Demo

*Please try it with us

iAQAR

Future works

1- Automate iAQAR
2- Enhance the performance

*Please try it with us

iAQAR

Thank you !

Any Questions?

iAQAR