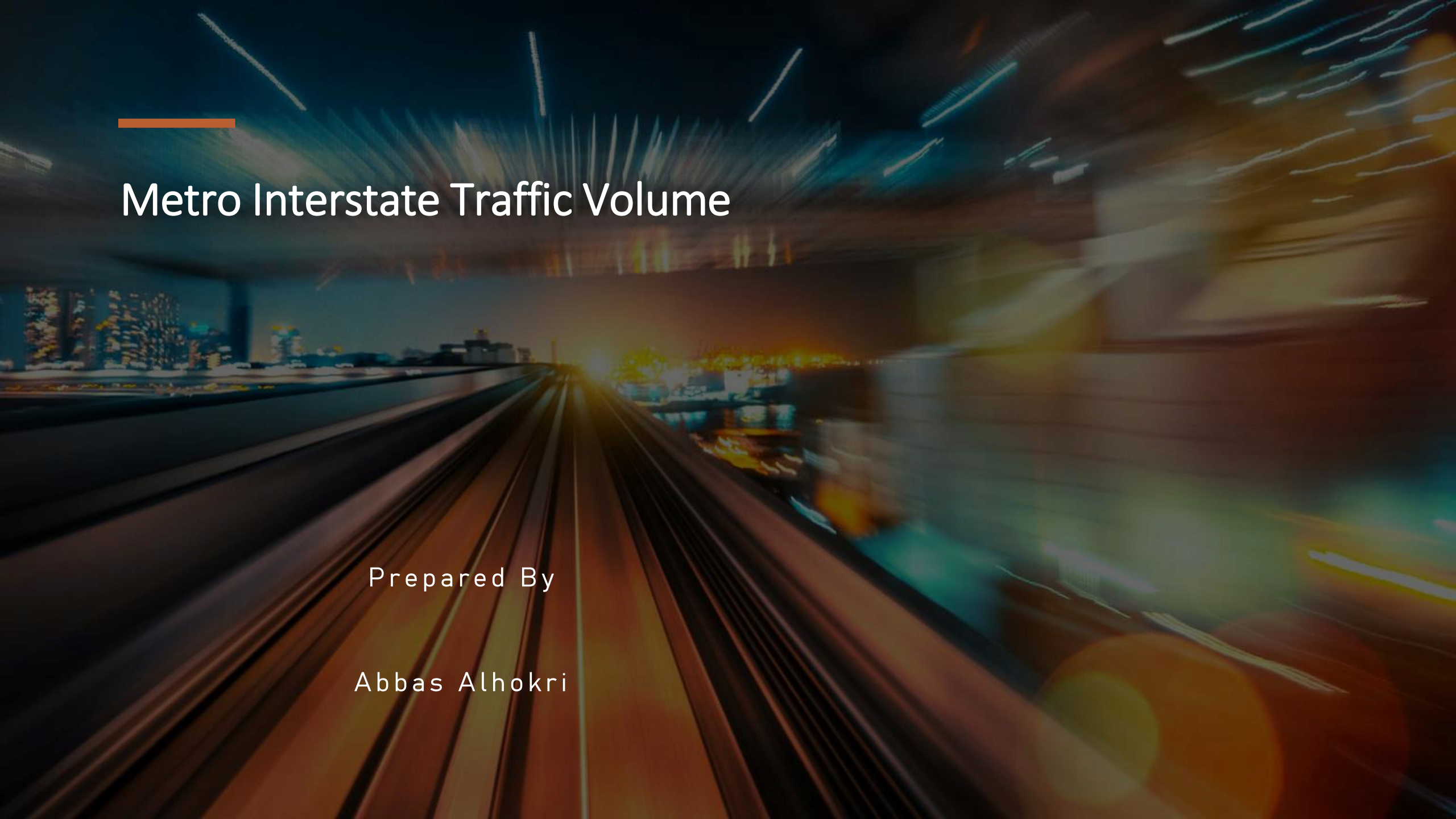


Metro Interstate Traffic Volume

Prepared By

Abbas Alhokri

Outline

- Problem statement
- Dataset description
- Data Preparation
- Feature selection
- Training models
- Result

Problem Statement

The purpose of this project is to Predict the volume traffic of the metro in Minneapolis state. In order to plan preventive maintenance for the metro and for advertisement purpose.

Dataset description

Feature	Data type	Example
holiday	Categorical	US National holidays plus regional holiday, Minnesota State Fair (e.g., Columbus Day , None)
temp	Numeric	Average temp (in Kelvin e.g., 288, 289)
rain_1h	Numeric	Amount in mm of rain (raining per hour - e.g., 0,1)
snow_1h	Numeric	Amount in mm of snow (snowing per hour - e.g., 0,1)
clouds_all	Numeric	Percentage of cloud cover (e.g., 45, 1)
weather_main	Categorical	Short textual description of the current weather (e.g, Clouds , clear)
weather_description	Categorical	Longer textual description of the current weather (e.g, sky is clear, few clouds)
date_time	DateTime	Date, Time and Hour of the data collected in local CST time (e.g., 2017-01-01 00:07:57)
traffic_volume	Numeric	Numeric Hourly I-94 ATR 301 reported westbound traffic volume (e.g., 6542,4516)

Data Preparation

- Data cleaning (e.g deleting outliers , drop null values)
- One hot encoding for weather_main column
- Split date_time into hour, day , month and year.
- Convert holiday column into Boolean valuue
- Creating day_off column following the logical function:
- off_day_colunm = Day + Holiday
- Plot Correlation heatmap to find the features related with target.

Feature selection

Correlated features with target
day_off
Temp
clouds_all
Hour
day_of_week

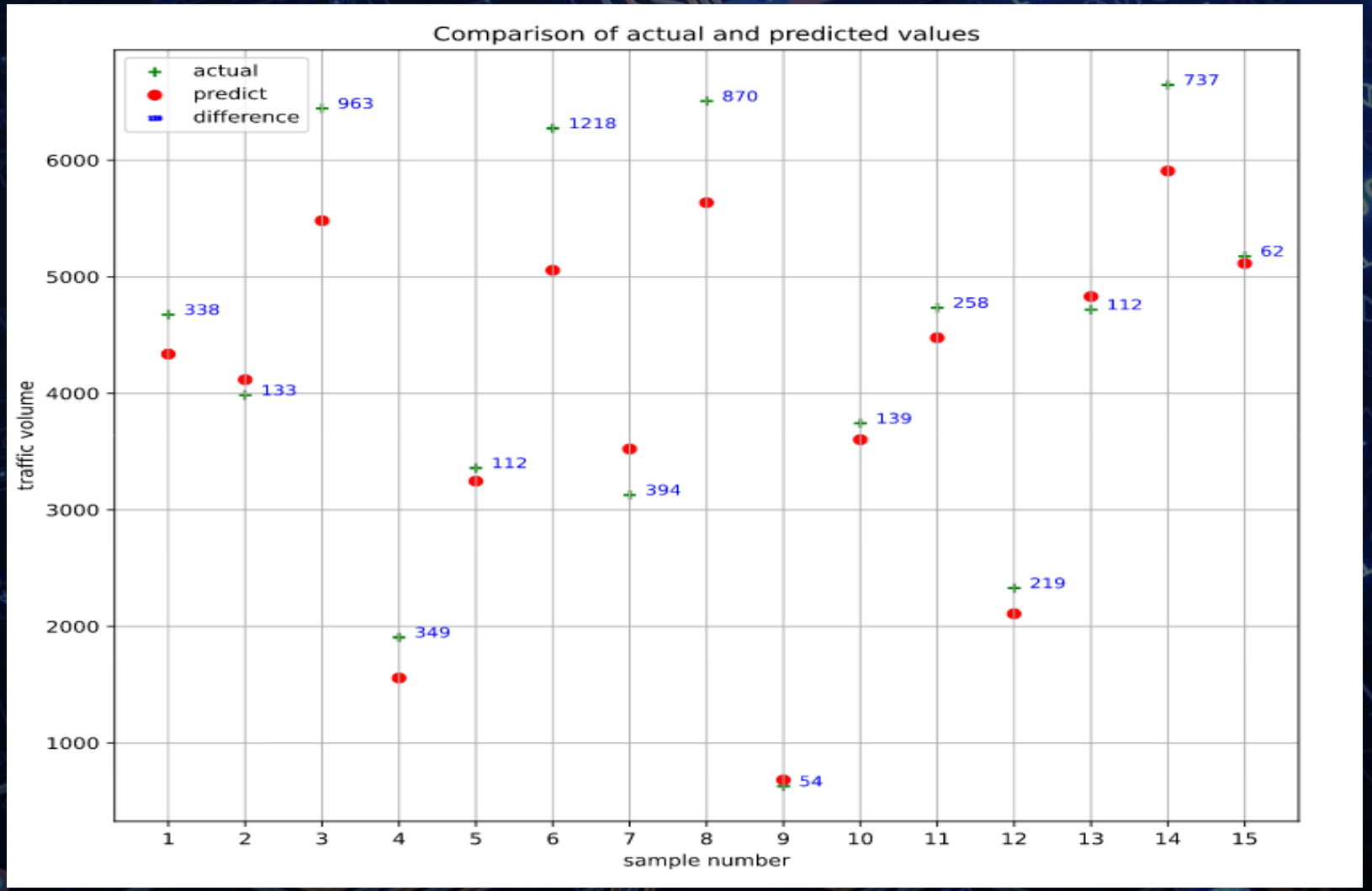
First result for three models on test set.

Model	R-sequare	MAE
Linear regression	18.2%	1583.8
Decision Tree Regressor	90.7%	389.46
Random forest Regressor	91.5%	363.98

Result after tunning parameters

Model	R-square	MAE	Best estimator
Decision Tree Regressor	90.7%	389.46	('max_depth': 12, 'min_samples_leaf': 50)
Random forest Regressor	91.5%	363.98	{'max_depth': 12, 'min_samples_leaf': 10, 'n_estimators': 75}

Visualization for the best models



Thanks For listening