Yiğit Yeldan
BSc METU Statistics
BSc Minor METU Computer Eng.
MSc METU Technology Policies
PhD... METU Technology Policies

Utku Civelek
BSc METU Electrical-Electronics Eng.
MSc METU Electrical-Electronics Eng.
MBA, METU
MSc IED Graphic Design
PhD... METU Informatics
Company Profile

With a turnover of 88.5 billion ₺
Turkey’s top industrial enterprise

With 5428 employees
The part of Koç Holding, the 35th best employer in the World

With 14.5 Nelson Complexity
One of the highest conversion rates in The Mediterranean Zone
An oil refinery produces value-added products such as gasoline, jet fuel, kerosene, diesel and LPG from crude oil. Uses the distillation process in which crude oil is heated at extreme temperatures to separate different hydrocarbons.
Refineries & Terminals

- İzmir
- İzmit
- Kırıkkale
- Mersin
- Antalya
- Giresun
- Kırıkkale
- Batum
- Marmara
- OPET
- 43,130 m³
- OPET
- 240,000 m³
- OPET
- 19,392 m³
- OPET
- 721,000 m³
Data Analytics

is the art of obtaining meaningful information

• to obtain unknown information from the current process
• better understand the past
• to make inferences for the future

with the help of software programs over data
Development of Data Analytics

Analytics 1.0  
90’s  
Descriptive Analytics  
• What happened?  
• Why?

Analytics 2.0  
2000’s  
Predictive Analytics  
• When?

Analytics 3.0  
Today  
Prescriptive Analytics  
• When and How?
Data Analytics in Refinery

- Improve Processes
- Safety & Security
- Increase Profitability

Qualified Human Resources
Cooperation with Technology Firms
Cooperation with Universities
Data Analytics Offices

ODTÜ·TEKNOKENT
Data Analytics Offices
Data Analytics Team

ODTU-TEKNOKENT

ITUARI
TEKNOKENT
Academic Collaborations
Data Infrastructure

99522 sensor
İzmit

38215 sensor
İzmir

11058 sensor
K.kale

5185 sensor
Batman
Data Infrastructure

640 million rows of data per day

İzmit
İzmir
K.kale
Batman
Data Infrastructure

640 million rows of data per day

Tupras Historian Database
Data Infrastructure

PHD

İzmit
İzmir
Kırıkkale
Batman

Data Storage
Technologies

Analytics

MAPR
DATA TECHNOLOGIES

OPEN TSDB

HADOOP

HONEYWELL
YOKOGAWA
Data Analytics Lifecycle

- Problem Definition
- Data Exploration
- Data Cleaning
- Modeling
- Monitoring
Data at Different Frequency

tmp = pd.resample(tseries, '1min', 'mean').interpolate
TwoStep = pd.resample(tmp, '50min', 'ffil')
Data at Different Frequency

Determination of the time interval with the highest correlation between them?

Lab Analysis Results - Relationship between Process Data?

Feature Engineering

Multicollinearity

PCA
Missing Values
Outliers

Isolation Forest
One Class SVM

Autoencoders,
Twitter/AnomalyDetection

Multivariate Outlier Detection
Bias-Variance Tradeoff

\[ \text{Err}(x) = \left( E[\hat{f}(x)] - f(x) \right)^2 + E \left[ \left( \hat{f}(x) - E[\hat{f}(x)] \right)^2 \right] + \sigma_e^2 \]

\[ \text{Err}(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error} \]
Model Validation

Validation Accuracy: 93%, 90%, 91%, 95%

Final Accuracy = Average(Round 1, Round 2, ...)

Nested Cross-Validation

Outer Loop
1. Train each split with optimal parameters
2. Average each split's test error

Inner Loop
Tune hyperparameters
Competencies for Working in the Field of Data Analytics

- Curiosity for data discovery
- Being a good listener to understand the problem
- Mathematics and statistics
- Competence in programming languages
- General information about big data management
- Data visualization skills
- Ability to transfer insights in a way that the customer can understand
Thanks

linkedin.com/company/tupras

linkedin.com/in/yigityeldan

linkedin.com/in/utkucivelek