EMRE OKCULAR

San Francisco, CA | 415 799 6601 | emreokcular@gmail.com | okcular.com.tr | github.com/emreokcular | linkedin.com/in/emreokcular

EDUCATION

University of San Francisco   San Francisco, CA
Master of Science in Data Science   August, 2020 – August, 2021(expected)
- Relevant Coursework: Python, Data Structures and Algorithms, Statistics, Data Visualization, SQL, Machine Learning, Deep Learning, NLP, Time Series, Product Analytics, Design of Experiments, Distributed Computing (Spark), Data Ethics

Yildiz Technical University   Istanbul, TURKEY
Bachelor of Engineering in Computational and Applied Mathematics   September, 2012 – June, 2017
- Relevant Coursework: Linear Algebra, Calculus, Discrete Mathematics, Abstract Mathematics, Algorithms, Programming, OOP, Probability, Statistics, Data Management, Optimization, Cryptography, Computer Vision, Image Processing, AI

Graz University of Technology   Graz, AUSTRIA
Erasmus Exchange Program in Computer Science and Mathematics   October, 2014 – August, 2015
- Ranked 1st among the outgoing Erasmus Exchange Program students in the department and awarded EU Scholarship.

WORK EXPERIENCE

Data Science Intern   January, 2021 – Present
Dictionary.com   Oakland, CA
- Predicting click-through rate (CTR) with random forest classifier trained on website logs and cookies. Increased model accuracy by 5% using sampling and feature engineering techniques in scikit-learn pipelines.
- Identifying most engaged user segments by exploring website behavior to gain insights for improving the ad auction.

Software Engineer   August, 2017 – October, 2020
Turkcell   Istanbul, TURKEY
- Achieved the ability to analyze streaming data in real-time and to take immediate actions through outbound communication channels by developing real-time streaming data analytics systems with Lean-Agile Methodologies.
- Expanded event-based scenarios such as gamification, anti-churn, up-sell, and retention resulting 15% increase in annual revenue by integrating high-volume data sources to complex event processing systems.
- Increased monthly bundle package sales by 15% discovering customer's opinions from inbound messages in SMS channels with NLP techniques such as named entity recognition, sentiment analysis, and text classification.
- Developed the capability of sending millions of messages and notifications per day for direct marketing team by building highly scalable campaign management applications with Java, PL/SQL, and Python.
- Empowered the marketing team to derive strategic insights for campaigns by creating a high-performance synchronous Java REST service for collecting and storing push notification responses in the Oracle SQL database.
- Automated the daily campaign report for all channels using a vast amount of relational data with PL/SQL procedures.

Data Science Intern   February, 2017– May, 2017
EVAM Streaming Analytics   Istanbul, TURKEY
- Explored streaming ML algorithm papers and identified algorithms to implement by understanding business needs.
- Enabled real-time clustering and outlier detection on streaming data by implementing density-based streaming clustering algorithm DenStream in the core product with Java.

DevOps Intern   June, 2016 – July, 2016
Anadolu Insurance   Istanbul, TURKEY
- Improved the application lifecycle by creating automated DevOps pipelines with build, test and deploy stages with Jenkins.

PROJECTS

- **Cancer Classification by Liquid Biopsy**

Achieved 73% accuracy with different models and placed in the top 10 in Kaggle private leaderboard. Fit and tuned various scikit-learn classifiers such as boosting and PyTorch deep neural networks to predict multi-class cancer types by identifying top features from microRNA profiles.

- **ML algorithm implementations from scratch in Python**

Regularized Linear and Logistic Regression with Gradient Descent, Naive Bayes, Decision Trees, Random Forest, K-means Clustering, Boosting, Deep Neural Networks, recommendation engine, and automated feature selection algorithms.

- **Ad Click Prediction**

Predicted ad clicks from KDD Cup dataset achieving 83% accuracy. Various classifier models, including tree-based ensemble methods used in scikit-learn pipelines applying feature selection, preprocessing, and tuning techniques with cross-validation.

TECHNICAL SKILLS

Python (Pandas, Scikit-learn, NumPy), PyTorch, Java, R, PL/SQL, NoSQL, Spark, Kafka, Linux, Git, Docker, AWS, Jenkins