EMRE OKCULAR

San Francisco, CA | 415 799 6601

emreokcular@gmail.com | **okcular.com.tr** | **github.com/emreokcular** | **linkedin.com/in/emreokcular**

WORK EXPERIENCE

Capgemini, San Francisco, CA **Oct, 2021 – Present**

Senior Data Scientist

- Created a hybrid demand prediction model combining clustering and curve-fitting models trained on billions of records using PySpark on Azure Databricks and improved store planning capabilities by optimizing shelf space of departments in full fleet.
- Built an end-to-end production-level ML pipeline, including auto feature selection, preprocessing, and evaluation stages.
- Crafted and optimized steps of the deployed ML pipeline that processes high volume delta tables with PySpark by code-level and cluster-level optimization techniques and reduced execution time by 20%.
- Deployed an MLOps process using MLFlow by registering the model as endpoints and serving the model as a REST API.
- Created neural networks to diagnose river blindness disease from biopsy images using PyTorch on AWS Sagemaker. Increased model accuracy by fine tuning pre-trained state-of-art object detection models and data augmentation techniques.

Dictionary.com, Oakland, CA

Data Science Intern **Jan, 2021 – Aug, 2021**

- Predicted click-through rate 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 within AWS (S3, EC2, EMR, RDS, Athena).
- Identified most engaged user segments by exploring website behavior to gain insights for improving the ad auction system.

Turkcell, Istanbul, TURKEY

Data Engineer **Aug, 2019 – Oct, 2020**

- Achieved the ability to analyze streaming data in real-time and take immediate actions through communication channels by developing big data processing systems with Java, Python and Agile Methodologies.
- Expanded event-based scenarios such as gamification, anti-churn, up-sell, and retention resulted in a 15% increase in annual revenue by integrating big data sources using Kafka, Spark, Logstash and SQL into complex event processing systems.
- Increased monthly bundle package sales by 10% discovering customers' opinions from messages in SMS channels with NLP techniques such as named entity recognition, sentiment analysis, and text classification using PyTorch NN.
- Performed affinity analysis on subscriber lifecycle events using apriori algorithm and increased campaign responses by 5% deriving insights for personalized bundle package campaigns.

Software Engineer **Aug, 2017 – Aug, 2019**

- Increased the daily capacity of sending millions of messages and notifications by 50% building highly scalable campaign management applications with Java, PL/SQL, and Python using best practices for software development lifecycle.
- Automated generation of daily campaign reports for all channels using a vast amount of relational data with tuned SQL queries and PL/SQL objects such as procedures, triggers, views, and tables.
- Empowered the marketing team to derive strategic insights for campaigns by creating a performant Java REST service for collecting push notification responses in the Oracle SQL database.

EDUCATION

University of San Francisco, San Francisco, CA

Master of Science in Data Science **Aug, 2020 – Aug, 2021**

- Coursework: Python, Data Structures and Algorithms, Statistics, Data Visualization, SQL, NoSQL, Machine Learning, Regression, Deep Learning, NLP, Time Series, Design of Experiments(A/B Testing), Distributed Computing (Spark)

Yildiz Technical University, Istanbul, TURKEY

Bachelor of Science in Computational and Applied Mathematics, *Honors Student* **Sep, 2012 – Jun, 2017**

Graz University of Technology, Graz, AUSTRIA

Erasmus Exchange Program in Computer Science and Mathematics, *Honors Student* **Oct, 2014 – Aug, 2015**

PROJECTS

- **ML algorithm implementations from scratch in Python** [Link]

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

- **Cancer Classification by Liquid Biopsy** [Link]

Achieved 73% accuracy 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, applying regularization techniques to predict multi-class cancer types.

- **Fabric Defect Detection with Deep Learning** [Link]

Predicted color, cut, thread, and hole defects on fabric with over 90% accuracy using image classifiers. Built model architectures such as CNN and ResNet18 using PyTorch by applying different image transformation and fine-tuning techniques.

TECHNICAL SKILLS

Python(Pandas, Scikit-learn, NumPy, SciPy), PyTorch, Java, SQL, NoSQL, Spark, Databricks, Kafka, Linux, Git, AWS, Azure