

# EMRE OKULAR

San Francisco, CA | 415 799 6601

[emreokcular@gmail.com](mailto:emreokcular@gmail.com) | [okcular.com.tr](http://okcular.com.tr) | [github.com/emreokcular](https://github.com/emreokcular) | [linkedin.com/in/emreokcular](https://linkedin.com/in/emreokcular)

## EDUCATION

**University of San Francisco**, San Francisco, CA

**Master of Science in Data Science**

**Aug, 2020 – Aug, 2021(expected)**

- 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**

- 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**

**Oct, 2014 – Aug, 2015**

- Ranked 1st among the outgoing Erasmus Exchange Program students in the department and awarded EU Scholarship.

## WORK EXPERIENCE

**Dictionary.com**, Oakland, CA

**Data Science Intern**

**Jan, 2021 – Present**

- 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.
- Performed exploratory analysis and forecasting on ad impressions to uncover the relationship between inner and ad clicks.

**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, 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.

## 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.

- **Ad Click Prediction** [\[Link\]](#)

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, SciPy, matplotlib), PyTorch, Java, SQL, NoSQL, Spark, Kafka, Linux, Git, Docker, AWS