

# 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** August, 2020 – August, 2021(expected)  
• Coursework: Python, Data Structures and Algorithms, Statistics, Data Visualization, SQL, 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** *Honored Student* September, 2012 – June, 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** 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  
• Predicted 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.  
• Identified most engaged user segments by exploring website behavior to gain insights for improving the ad auction.  
• Performed exploratory analysis and forecasting on ad impressions to uncover the relationship between inner and ad clicks.
- Software Engineer** August, 2017 – October, 2020  
**Turkcell** Istanbul, TURKEY  
• Achieved the ability to analyze streaming data in real-time and take immediate actions through communication channels by developing big data processing systems with Lean-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 into complex event processing systems.  
• Increased monthly bundle package sales by 15% discovering customers' opinions from 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 by building highly scalable campaign management applications with Java, PL/SQL, and Python using best practices for software development lifecycle.  
• 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.  
• Automated generation of daily campaign reports for all channels using a vast amount of relational data with tuned SQL queries and PL/SQL procedures.
- Data Science Intern** February, 2017 – May, 2017  
**EVAM Streaming Analytics** Istanbul, TURKEY  
• Explored streaming ML algorithm papers and identified models to implement by testing and understanding business needs.  
• Enabled real-time streaming clustering and anomaly detection by implementing DenStream clustering algorithm with Java.

## 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