Emin Can Oguz

Istanbul/Turkey
emincanoguz11@gmail.com
emincanoguz11.github.io

ABOUT ME

As a data engineer with 3 years of experience, I specialize in designing, building, and maintaining high-performance data systems. I have experience managing large-scale data pipelines, implementing streaming solutions, and optimizing costs and queries within clusters. In particular, I have successfully managed data infrastructure for multiple Black Friday events, delivering reliable performance and insights to stakeholders.

EDUCATİON

Maltepe University- Software Engineering (2016-2021)

LANGUAGES

English - Fluent
German - Intermediate
Russian - Intermediate

WORK EXPERİENCE

Hepsiburada - Data Engineer (April 2021 - Present)

- Developed a Data Mesh pipeline for all teams, enabling efficient data sharing and collaboration.
- Implemented a Data pipeline to process and transform large-scale datasets for analytics and business intelligence purposes.
- Utilized real-time technologies to create a streaming job, enabling the processing of data in real-time.
- Successfully managed a cluster exceeding 10PB, ensuring high availability and scalability of the data infrastructure.
- Created Airflow pipelines to orchestrate and schedule data workflows, improving efficiency and reliability.
- Focused on optimizing costs and queries within the data cluster, enhancing performance and reducing expenses.

Doctorla aka Online Saglikcim - Junior Data Scientist (April 2020 –Jan 2021)

- Developed and deployed NLP models using Tensorflow for analyzing and extracting insights from healthcare data.
- Leveraged machine learning techniques to solve real-world problems in the healthcare domain.
- Worked with large-scale datasets and applied advanced techniques to extract meaningful insights.
- Created predictive models to support decision-making processes and improve healthcare outcomes.

RemoteTeam.com - Junior Data Scientist (May 2020 - Jan 2021)

- Utilized recommendation system algorithms to generate personalized recommendations for users.
- Developed and deployed recommendation algorithms to enhance user experience and increase engagement.
- Worked with large-scale datasets to identify patterns and trends, improving the accuracy of recommendations.
- Utilized advanced data analysis techniques to extract insights and optimize recommendation performance.
- Conducted A/B testing to evaluate the performance of recommendation algorithms.
- Collaborated with software engineers to integrate recommendation models into production systems.
- Contributed to the development of data-driven strategies to improve user engagement and retention.

Yesil Science - Intern (April 2020 – June 2020)

- Developed a Mandarin project for scoring healthcare using Machine Learning techniques.
- Leveraged machine learning algorithms to analyze Mandarin healthcare data and derive meaningful insights.
- Worked closely with a team of researchers to develop and refine the scoring model.
- Gained hands-on experience in data preprocessing, feature engineering, and model evaluation.

SKILLS

- Strong knowledge of data engineering principles and best practices.
- Proficient in programming languages such as Python, Java, and SQL.
- Experience with big data technologies like Hadoop, Spark, and Kafka.
- Familiarity with cloud platforms such as AWS or GCP.
- Experience with data processing frameworks like Apache Airflow.
- Solid understanding of machine learning algorithms and statistical analysis.
- Strong problem-solving and analytical skills.
- Excellent communication and collaboration abilities.

TECHNOLOGIES

Python, Java, ,Scala, C#, SQL, NoSQL, Cloud Technologies (AWS,Google Cloud), Prometheus, Apache Druid, Apache Spark, Apache Hadoop, Apache Kafka, RabbitMQ, Linux, Kubernetes, Docker, Gitlab CI/CD, Jenkins, Tensorflow, Spark NLP, Scikit-Learn, Google BigQuery, Google Cloud Storage. Apache Airflow, DBT, Amundsen. Datahub, MongoDB