ALI EMRE VAROL

+90(542) 657-6834 · a.emrevarol@gmail.com · LinkedIn · GitHub · Medium · Istanbul, Turkey

SUMMARY

As a Senior Data Scientist, I integrate advanced analytics, machine learning, and NLP to drive transformative solutions. My expertise lies in developing AI-driven platforms, improving decision-making processes, and leading data science projects from concept to completion. I am committed to mentoring emerging talent and am currently advancing my knowledge through a Ph.D. in Industrial Engineering. I leverage the latest technologies to address complex business challenges, enhance operational efficiency, and support strategic growth initiatives. My approach is centered on creating impactful and innovative solutions that effectively utilize data to inform business strategies.

EDUCATION

Middle East Technical University, *Ankara, Turkey* — Expected Graduation date: 2024
Degree: Ph.D. in Industrial Engineering — GPA: 3.4 / 4.0
Thesis: Vehicle Routing solutions for Unmanned Aerial Vehicles

Science Institute of Turkish Naval Academy, Istanbul, Turkey — 2012
Degree: Master of Science in Industrial Engineering — GPA: 3.77 / 4.0
Thesis: Simulation Modeling of Maritime Piracy using Discrete Event and Agent-Based Approaches

Turkish Naval Academy, *Istanbul, Turkey* — 2005
Degree: Bachelor of Technology in Industrial Engineering — GPA: 3.9 / 4.0
Achievements & Awards: **Distinguished graduation, 2nd** out of 240.

EXPERIENCE

Berlin Tires, Berlin, Germany (Remote) — February 2023 – Present
Senior Data Scientist

- **Led the generation of advanced forecasting models** that revolutionized inventory management across Production & Sales, resulting in a significant reduction of excess stock and cost savings by accurately predicting demand.
- **Developed dynamic, real-time dashboards** for various departments, providing critical insights that enhanced data-driven decisions and fostered collaborative efforts, significantly improving operational efficiency.
- **Pioneered an AI-powered Q&A assistant** for Sales and Customer Support, boosting self-service capabilities and streamlining interactions, which led to faster query resolution and improved customer satisfaction.
- **Executed in-depth analytics on application data**, uncovering key insights that drove strategic business decisions, optimized processes, and enhanced customer engagement through a data-centric approach.
- **Overhauled 25 years of historical data**, ensuring integrity and reliability through rigorous cleaning, setting a solid foundation for future analytics and strategic decision-making.
- **Achieved a groundbreaking 95% accuracy rate in quarterly forecasting models**, significantly surpassing previous projections by 30% and setting a new standard for forecasting precision within the company.
  Advanced Skills: FastAPI, PostgreSQL, Darts, Nixtla, FB Prophet, Streamlit, Docker, Pandas, NumPy, DuckDB, Generative AI, GPTs, LangChain, LlamaIndex, RAG

Eyva, Cologne, Germany (Remote) — June 2022 – December 2022
Senior Data Scientist

- **Developed NLP-based Voice of Customer models** to extract critical insights, significantly improving understanding of customer needs and preferences.
- **Created a Concept Model with Neo4j Knowledge Graph** to refine Beauty Product strategies, enabling more targeted marketing and product development efforts.
- **Utilized Darts and Nixtla libraries to develop advanced forecasting models** for Beauty Products and search terms, driving strategic decisions with accurate sales predictions.
- **Implemented Sentence BERT Transformers in Classification Models** to optimize search term relevancy, markedly improving user search experience.
- **Customized sentiment analysis models using the SetFit approach**, achieving a 30% improvement in precision, recall, and accuracy, thereby enhancing customer feedback analysis.
  Advanced Skills: FastAPI, Neo4j, Darts, Nixtla MLForecast, Streamlit, Docker, HF Transformers, S-BERT Transformers, Pandas, NumPy, DuckDB, SpaCy

John Snow Labs, Lewes, DE USA (Remote)   Aug 2021 – June 2022
Senior Data Scientist

- **Spearheaded the development and deployment of transformative NLP and AI solutions**, serving as proofs of concept and technological advancements for leading health groups, hospitals, and pharmaceutical companies, including Optum Health, CVS Health, Philips, Veradigm, Boehringer Ingelheim, and Onco Health. These solutions have catalyzed data-driven decision-making and enhanced patient outcomes.
- **Pioneered the application of advanced NLP techniques** to build real-time healthcare solutions, harnessing the power of BERT models to process complex medical reports, perform identification and de-identification, generate medical text using GPT family models, and establish meaningful connections between medical concepts. TensorFlow, PyTorch, and PySpark have been pivotal tools in this endeavor.
- **Collaborated seamlessly with cross-functional teams to address diverse NLP use cases**, ranging from semantic analysis to sentiment extraction, effectively bridging the gap between technical capabilities and healthcare domain requirements.
- Developed a sophisticated medical question-and-answer application through the implementation of medical knowledge graphs, facilitating streamlined access to critical medical insights.
- **Engineered a FastAPI-based solution to serve Spark NLP as an AutoNLP platform**, revolutionizing Natural Language Understanding tasks and enabling rapid deployment of language models.
- **Excelled in data analysis and visualization,** employing the ML tech stack including Pandas, NumPy, Seaborn, Plotly, and Streamlit, alongside Neo4j-powered knowledge graphs, to distill actionable insights from complex healthcare data.
- **Demonstrated expertise in image processing by utilizing OpenCV for Spark OCR,** coupled with official training responsibilities as a Spark OCR trainer in certification programs.
- **Proficiently conducted text preprocessing, entity recognition, and relation extraction** utilizing a range of tools, including NLTK, Transformers, SpaCy, and Spark NLP.
- **Mastered various Transformer models** such as BERT, Roberta, Distilled versions, and pre-trained clinical embeddings, including BioBERT and T5, enhancing healthcare text processing capabilities and driving efficient training and fine-tuning, including TrOCR tasks.
- **Championed MLOps principles** by implementing model serving and experiment tracking using SynapseML, FastAPI, and MLFlow within Databricks environments, ensuring seamless model deployment and continuous improvement.
- **Leveraged an in-depth understanding of healthcare dynamics** to excel in sales engineering across Healthcare, Pharma, Payer, and Provider segments, effectively bridging the gap between technological innovation and practical application.
- **Partnered with stakeholders to design tailored solutions** in line with healthcare sector standards.
  Advanced Skills: **FastAPI, Neo4j, PySpark, Streamlit, Docker, HF Transformers, SynapseML, TensorFlow, Keras, Pandas, NumPy, Plotly, PyTorch, BERT-based LMs, GPT-family text generation models**

Two Desperados, Belgrade, Serbia (Remote)   Mar 2020 – Feb 2021
Data Scientist

- **Engineered and implemented sophisticated machine learning models** to predict user churn, segment customers, forecast lifetime values, and curate personalized offers for video game enthusiasts. This comprehensive approach led to an impressive 40% increase in day-7 retention rates and a 10% surge in in-app purchases, driving significant revenue gains.
- **Designed and crafted analytic dashboards** utilizing Plotly, Plotly Dash, and Google Data Studio, providing real-time visualizations of key performance metrics. These interactive dashboards empowered stakeholders to make informed decisions and fine-tune strategies for optimal player engagement.
- **Established a robust artificial intelligence architecture**, including efficient data pipelines, to streamline data processing and model deployment. This foundation paved the way for swift, scalable, and efficient integration of machine learning solutions into the gaming ecosystem.
- **Collaborated seamlessly with the user acquisition team** to optimize ad campaigns, leveraging ROI/ROAS analyses to maximize ad spend efficiency and elevate user acquisition efforts.
- **Played a pivotal role in cross-functional collaboration**, working closely with User Acquisition, Monetization, and Product Teams to align data-driven strategies with broader organizational goals. This synergy enabled holistic player experiences and revenue enhancement.
  Advanced Skills: **Google Cloud Platform Vertex AI, Google Data Studio, Plotly Dash, Python, Scikit Learn, TensorFlow, Keras, XGBoost, BigQuery, TensorFlow Extended (TFX), A/B Testing**

Datajarlabs, Ankara, Turkey (Remote)   Apr 2019 – Jul 2020
Data Science Consultant and Mentor

- **Spearheaded machine learning projects**, taking the lead in model development and optimization to deliver data-driven solutions that address complex business challenges. My hands-on approach to ML has consistently yielded actionable insights and boosted organizational performance.
- **Leveraged my expertise to mentor and guide aspiring data scientists** in their journeys, offering tailored support and fostering a culture of continuous learning. As a data science tutor, I have nurtured emerging talents, equipping them with the tools to excel in the ever-evolving data landscape.
- **Provided invaluable consultation on data science projects**, collaborating closely with teams to identify opportunities for leveraging data for strategic decision-making. My insights have driven data-driven initiatives, leading to enhanced operational efficiency and increased competitiveness.
- **Contributed to data science training bootcamps as a mentor**, imparting essential skills and best practices to future data science professionals. My dedication to mentorship extends beyond the classroom, as I take pride in guiding individuals toward success in their careers.
- **Presented captivating lectures on data science**, with a particular focus on machine learning, captivating audiences with my expertise and real-world case studies. My engaging teaching style has fostered a deep understanding of complex concepts and inspired a new generation of data enthusiasts.
- **Led comprehensive statistical analyses**, providing data-driven recommendations and actionable insights that have steered critical business decisions.
- **Managed a data scientist hiring Datathon** for "Yildiz Holding", orchestrating a competitive environment where top talent was identified and recruited to strengthen the company's data science capabilities.
  Advanced Skills: Python, Scikit Learn, TensorFlow, Keras, Pandas, NumPy, Plotly, Plotly Dash, Flask

Government, Turkey   Aug 2005 – Mar 2019
Data Analyst & Operations Research Specialist

- **Developed and successfully deployed an enterprise-level decision support system**, leveraging industrial engineering and operations research techniques. This comprehensive solution revolutionized critical aspects of the organization, including cost-effectiveness, stock planning, workforce optimization, and scenario-based comparisons. The system's impact extended far beyond its initial scope, fueling strategic insights and informed decision-making.
- **Orchestrated meticulous analyses and planning for inventory management**, playing a pivotal role in optimizing inventory levels, minimizing costs, and ensuring timely availability of resources. My data-driven approach enabled the organization to balance supply and demand with precision.
- **Crafted and presented analytic dashboards**, offering intuitive visualizations of key performance indicators to C-level managers. These dashboards provided real-time insights, facilitating informed decision-making at the highest levels of the organization.
- **Distinguished as a key member of the Project Management Board,** I assumed the leadership role of the Computer Software Test Group. This multifaceted responsibility involved coordinating testing efforts, ensuring software quality, and orchestrating seamless project execution.
- **Led the preparation of software requirement analyses and the design** of computer software units, collaborating with cross-functional teams to translate complex technical requirements into practical, user-centric solutions.
- **Successfully navigated "Stealth Projects",** demonstrating adaptability and tenacity in handling confidential initiatives and driving their successful execution.
  Advanced Skills: Java, Spring, Hibernate, Oracle DB, Linux

TECH SKILLS

Data Science	:	TensorFlow, XGBoost, Light GBM, Keras, Spark NLP, PyTorch, HF Transformers, Scikit Learn Algorithms, FastAPI, TensorFlow Extended, Darts, Nixtla StatsForecast, Nixtla ML Forecast, Pandas, NumPy, Matplotlib, Seaborn, Plotly, Plotly Dash, PySpark, FastAPI, Flask, Streamlit, Spark OCR, Spark NLP, Spark NLP HealthCare
Generative AI	:	LangChain, LlamaIndex, RAG, GPTs
Cloud Technologies	:	Google Cloud Platform (GCP) Vertex AI, GCP Looker Studio, Databricks
Databases	:	BigQuery, PostgreSQL, Oracle, Noe4j Graph DB, DuckDB
Others	:	Git, Docker, Linux, Postman, Jira, Trello, Tableau

PUBLICATIONS

- **Understanding COVID-19 News Coverage using Medical NLP**, 44th European Conference on Information Retrieval (ECIR'22) in Stavanger, Norway · Mar 19, 2022
- **Simulation Prevention Operations at Sea against Maritime Piracy**, Journal of Operational Research Society · Dec 1, 2015
- **Simulation Modeling of Maritime Piracy using Discrete Event and Agent-Based Approaches**, Proceedings of the 3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications · Jan 1, 2013

BLOG POSTS

- **Creating Clinical Knowledge Graph by Spark NLP & Neo4j,** Medium, 2021
- **Exploring the European Natural Gas Network as a Knowledge Graph,** Medium, 2022
- **Analyzing Roland Garros and US Open Tennis Tournaments Via Neo4j,** Medium, 2022

CERTIFICATES

- Noe4j Certified Professional
- Spark NLP Data Scientist
- Spark NLP Healthcare Data Scientist
- Google Cloud Platform Big Data and Machine Learning Fundamentals
- Sequences, Time Series and Prediction
- End-to-end Machine Learning with TensorFlow on GCP
- How Google does Machine Learning
- Unsupervised Machine Learning for Customer Market Segmentation

REFERENCES

- References available upon request.