SAHAR KHALAFI

Phone number: (+98) 9352655817 (Mobile)
Email address: sahar.khalafi66@gmail.com
LinkedIn: https://www.linkedin.com/in/sahar-khalafi-473631176
GitHub: https://github.com/saharkhalafi
Date of birth: 18/11/1995 | Nationality: Iranian | Gender: Female

ABOUT ME

Data scientist with over 6 years of experience in machine learning, deep learning, and data analysis, specializing in text mining, natural language processing, and predictive modeling. Experienced in developing automated systems for healthcare and e-commerce applications, including phenotype identification, clinical text summarization, and recommendation systems. Skilled in Python, R, SQL, and deep learning frameworks, with a strong track record of designing and implementing AI-powered solutions.

EDUCATION

2018 – 2020 Isfahan, Iran
MASTER'S IN SOFTWARE ENGINEERING Isfahan University of Technology (IUT)
Specialized in text mining and machine learning for medical texts. Developed a phenotype-based medical text summarization system using deep learning and knowledge base integration to automatically extract semantic features from clinical notes, improving F1-score over conventional methods. Published a Q1 journal paper on phenotype prediction from clinical notes.
Thesis: "Phenotype-Based Summarization of Electronic Health Records" [Link]

2013 – 2017 Alborz, Iran
Bachelor's IN SOFTWARE ENGINEERING Karaj University

PUBLICATIONS

Deep Learning Approach for Phenotype Prediction from Clinical Notes (2021)
Authors: Sahar Khalafi, Nasser Ghadiri, Milad Moradi
Journal of Ambient Intelligence and Humanized Computing [Link]

Designing an Improved Deep Learning-Based Classifier for Breast Cancer Identification in Histopathology Images (2022)
Authors: Amirreza Babaahmadi, Sahar Khalafi, Fatemeh Malekipour
International Conference on Machine Vision and Image Processing (MVIP), IEEE [Link]

Designing an Improved Deep Learning-based Model for COVID-19 Recognition in Chest X-ray Images: A Knowledge Distillation Approach (2023)
Authors: Amirreza Babaahmadi, Sahar Khalafi, Masoud Shariatpanahi, Moosa Ayati
Revised, Iran Journal of Computer Science/Springer Journal [Link]

DIGITAL & TECHNICAL SKILLS

Machine Learning & AI
- Clustering, Regression, Classification
- Natural Language Processing (NLP)
- Deep Learning: Transformers, Keras, OpenCV
- Model Development & Evaluation

Programming & Tools
- Python, R, SQL (MySQL / SQL Server)
- Data Visualization: Power BI, Google Data Studio
- Jupyter Notebook
- Scikit-learn, NumPy, Pandas, Seaborn

Business Intelligence & Analytics
- Customer Segmentation
- ETL Pipelines & Data Processing
- Data Warehousing

PROJECTS

Medical Text Analysis & Disease Prediction

- Designed NLP-based models to extract clinical insights and predict diseases from unstructured medical text.
- Tools: Python, Keras, BioBERT, ClinicalBERT

Personalized Medical Text Summarization

- Built an abstractive summarization system tailored to individual physician reading patterns using transformer-based models.
- Tools: ClinicalBERT, textual similarity, ontology matching, Python

Customer Next-Basket Prediction & Purchase Pattern Mining

- Developed predictive models to forecast next purchases and uncover buying patterns using transaction data.
- Tools: Python, scikit-learn, XGBoost, MLP, Association Rule Mining, Apriori

Automated Product Image Enhancement for E-Commerce

- Created a deep learning pipeline for auto-retouching and standardizing product images to improve visual consistency.
- Tools: Python, OpenCV, U2Net, Image Augmentation

Hybrid Product Recommendation System

- Proposed and implemented a hybrid recommendation engine combining NLP-based similarity, association rules, sales data, and gender filter
- Tools: Matrix Factorization, NLP, Association Rules, Python

Product & Customer Segmentation

- Applied clustering and classification techniques to segment users and group products based on purchase behavior and textual similarity.
- Tools: K-Means, Hierarchical Clustering, PCA, RFM, NLP, Python

CERTIFICATIONS

Supervised Machine Learning: Regression and Classification
Coursera (offered by DeepLearning.AI) – Completed January 2024
33 hours | Verified Certificate

Machine Learning with Python
Coursera (offered by IBM) – Completed December 2023
20 hours | Verified Certificate

WORK EXPERIENCE

09/2023 – CURRENT Tehran,Iran
MARKETING DATA SCIENTIST GOLESTAN CO

- Conducted in-depth data analysis and applied machine learning techniques to build predictive and statistical models, including a GRU-based next-basket forecasting model (65% accuracy) and KMeans clustering for customer segmentation (10% improvement in pattern identification).
- Built ETL pipelines and automated data workflows using Python and OpenCV, increasing processing capacity from 100 to 3,000 items per day.
- Developed AI-based solutions to anticipate patterns, segment behavior, and support data-driven decision-making.

09/2021 – 08/2023 Tehran, Iran
MARKETING DATA ANALYST MODISEH

- Designed and implemented interactive dashboards and visualizations to communicate insights from predictive and machine learning models to stakeholders.
- Conducted quantitative analysis of trends and KPIs, translating complex datasets into actionable recommendations and supporting data-driven research decisions.

LANGUAGE SKILLS

- Persian: Native
- English: Intermediate level (currently improving; official certification in progress)
- Turkish: Intermediate (B2)