
Summary of Qualifications

As a self-learner data scientist and researcher i can help to transform data to business insights by formulating business problems to data analytics use cases and projects , providing solutions by statistical and machine learning modeling, extract hidden patterns, forecast future events, detect anomalies ,recommend your customers best products and services and so on, finally propose data driven solutions at scale and deliver business value. During the last eight years, i have worked on sales and marketing, banking, insurance, manufacturing, health care ,e-business and transportation business problems with more than one hundred real world data sets. Also i have more than ten years experience in software modeling and development , as programmer, system analyst and methodologist and project manager.

ML tasks i worked on the most, tabular(iid and non-iid) , text and audio:

- Time series analysis and forecasting, classification and anomaly detection providing competitive and modern effective solutions for popular business use cases
- Recommendation systems development(e.g. Job Recommendation)
- Predictive modeling, algorithms and techniques,tree-based, linear, gbms, ensembling, deep and transfer learning and self-supervised learning applying many python open source libraries.
- NLP and Text Analytics, deep and transfer learning
- XAI/MLI , ML models interpretation for real world problems(e.g. SHAP framework, N-BEATS)
- Distributed ML and Cloud : Azure Machine Learning & MLOps , Azure Databricks , Azure dsvm, TensorFlow , pySpark((ML, MLlib), RAPIDS

Education

- *BSc. in Mathematics and Computer Science - Sharif University of Technology , 1996*
- *MSc. in Artificial Intelligence - Amirkabir University , 2001*

Licenses & certifications

- [Recommendation Systems on Google Cloud](#), Google GCP
- [Introduction to Generative AI](#) , Google GCP
- [Introduction to Large Language Models](#) , Google GCP
- [Kaggle Competitions Master\(2018\)](#), Kaggle
- [Kaggle Dataset Master\(2021\)](#) , Kaggle

Communities and Awards

Seven years research and participation in kaggle([Profile](#))



[Kaggle Competitions Master\(2018\).](#)

[Kaggle Dataset Master\(2021\)](#)

[Code and Discussion Expert\(2019\)](#)

[KaggleX BIPOC Mentor \(2023\)](#)

Kaggle Competitions highlights :

- *TReNDS Neuroimaging , Gold medal(6th of 1047). Multiscanner normative age and assessments prediction with brain function, structure, and connectivity.*
- *IEEE-CIS Fraud Detection , Gold medal (8th of 6351). Detecting fraud from customer transactions .*
- *LANL Earthquake Prediction ,Silver medal. Predicting upcoming laboratory earthquakes.*
- *Elo Merchant Category Recommendation , Silver medal. Helping understand customer loyalty with the existing recommender system.*
- *Jane Street Market Prediction, Quora Insincere Questions(NLP, Detect toxic content to improve online conversations Classification) and American Express Default Prediction (credit default prediction) , 3 Bronze medals*

Consultancy and Projects

Cinere, Consultant , 2019

[Sales and Marketing](#)

Customer analytics use cases including:

- *Customer segmentation, RFM,CLV prediction*
- *Sale Forecast per channel and product category, multivariate , multi-step **time series forecasting**.*
- *Up-Lift for promotions prediction, multiclass, targeting optimization(base) , attribution modeling (MCMC pymc) and deliver operational solutions for business value.*

Azki/Bimito, Consultant, 2020

[Insurance Recommendation algorithms benchmark](#)

*Tabular data , full pipeline preprocessing and transformation, **recommender algorithms**, Matrix factorization, collaborative filtering(classic vs VAE) , content based, GBM-based, NCF*

AutojobApply, 2023 (remote project)

[Job recommendation system](#)

- *Elasticsearch 8.x query design and embedding to site backend, geo-distance, **full_text search** queries, semantic search*
- *Applying benchmark on **cold start** algorithms on phase (I) of interaction data(e.g. LightFM)*
- *Research on recommendation algorithms and survey development, collaborating filtering , content based and hybrid algorithms*

Positions

KeyLeadhealth

Melbourne , Australia

Senior Data Scientist and
Machine Learning
Engineer

2020 - Present

Health care, tabular, text, audio/signal ML use cases, applying Azure products and services e.g. Azure Machine Learning & MLOps , Azure Databricks , Azure dsvm.

- ICU and hospital readmission , mortality, phenotype and LOS prediction, dynamic , single and multistep(**Direct, DirRec, Rectify**), **time series forecasting** and classification. Keras RNNs, NBEATSModel, NHiTSModel , XGB, and more.
- **Transfer Learning** , applying on clinical notes, Bert ,**ClinicalBert**, Huggingface for clinical trials and readmission and mortality prediction
- Drug Repurposing solutions on RWD and EHR data (Recommendation algorithms)
- Covid-19 detection using acoustic breathing/cough/speech signals (PASE+ , VGGish, **Self-Supervised** Pre-Training wav2vec2, benchmarking)

Saman Bank

Data Science Team lead

2018 - 2019

Responsible for manage phase I of Blue **Bank data** analytics solutions:

- Team building and team mentoring & training
- **Predictive modeling** use cases formulation(payment, loan, Insurance) and test pipeline algorithms implementation at scale using Big Data stack: pySpark(ML-MLlib), Microsoft Machine Learning for Spark MML
- Business analytics and Big Data transition proposal development

Mellat Bank

Data Scientist

2016-2017

- **Marketing** and transactional anomaly detection ML use cases implementation on Big Data Stack (Hadoop, pySpark)
- **Customer analytics** ML use cases implementation and deployment on Big Data Stack (pySpark), a variety of tree based and linear models and advanced anomaly detection algorithms.
- Cloudera Hadoop, Sqoop, Impala, pySpark Big Data Stack

MIRAS Technologies

ML Engineer

2015-2016

Focusing on **Text Analytics**:

- Samsung , NLP , effects of news on **marketing**
 - Full crawling framework implementation by Scrapy
 - Text analytics, classification, sentiment analysis, entity extraction, Regex, NLTK, Gensim, Glove ,Word2Vec, by: Keras and Tensorflow, Scrapy, ElasticSearch
 - Generate pre-trained models for persian language
- Digikala text analytics , product segmentation and classification , MongoDB, topic modeling, next word prediction, summarizing.

Fanap

Data Analyst

2014-2015

- MIDCO project , sensor data MES (**Manufacturing**) and supply chain data analysis(Batch)
- KPI definition and analysis(Balanced Scorecard) and bridging KPI framework to CRASP

Tahlilgaran novin

Tehran, Iran

Data Warehouse Designer

2011-2012

- Ministry of Transportation and Urban Development project , **Data Warehouse** design (Kimball method and patterns in dimensional modeling) and data driven business process analysis.