Ali Janalizadeh Choobbasti

Unit 34, Block 2, Negin Residence, Doctor Kariman Street, Haj Poor Amir Street, Sazman Ab Alley, Sheikh Fazlollah Noori Highway, Tehran, Iran.
ali.janalizadeh@outlook.com +98 938 367 3878 https://ajanaliz.github.io/

RESEARCH INTERESTS

- Reinforcement Learning
- Online Learning
- Adversarial Learning
- Multi-Agent Systems
- Natural Language Processing

AREAS OF INTEREST

- Stock Market Prediction
- Dynamic Pricing
- Game A.I.
- Dialogue Systems

EDUCATION

- **Amirkabir University of Technology (Tehran Polytechnics)** Sep 2014 – Present
  (**Ranked 80** globally by subject of Engineering)
  - –B.Sc. in Computer Engineering - Information Technology
    - –Cumulative GPA (via 120 passed units): 17.11/20
    - –**Last 4 semesters' GPA: 17.74/20**
    - –**Selected Coursework GPA*: 18.99/20**
- **Shahed Highschool** Sep 2010 –Jun 2014
  - –High School Diploma
    - –Field: Mathematics and Physics
    - –Cumulative GPA: 19.07/20

PUBLICATIONS

- **Advertisement Recognition Using Mode Voting Acoustic Fingerprint**, Reza Fahmi, Hosein Abedi Firouzjaee, **Ali Janalizadeh Choobbasti**, S. H. E. Mortazavi Najafabadi, Saeid Safavi, International Conference on Robotics and Machine Vision (ICRMV 2017), vol. 10613, p. 106130D. International Society for Optics and Photonics, Dec 2017
- **MirasVoice A bilingual (English-Persian) speech corpus**, Amir Vaheb, **Ali Janalizadeh Choobbasti**, S. H. E. Mortazavi Najafabadi, Saeid Safavi, Behnam Sabeti, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), May 2018
- **MirasText An Automatically Generated Text Corpus for Persian**, Behnam Sabeti, Hossein Abedi Firouzjaee, **Ali Janalizadeh Choobbasti**,S. H. E. Mortazavi Najafabadi, Amir Vaheb, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), May 2018

PAPERS IN PREPERATION

- **Investigating Language Variability on the Performance of Speaker Verification Systems**, Amir Vaheb, **Ali Janalizadeh Choobbasti**, S. H. E. Mortazavi Najafabadi, Saeid Safavi
  This paper has been submitted at the International Conference on Speech and Computer (SPECOM 2018)
- **MirasSentiment The Largest Labeled Text Corpus for Sentiment Analysis**, Behnam Sabeti, **Ali Janalizadeh Choobbasti**, S. H. E. Mortazavi Najafabadi. Feb 2018 – Present
  In this project we have gathered the largest comment corpus for sentiment analysis on Farsi text. This corpus includes **900,000** automatic labeled comments and over **100,000** crowd-sourced labeled comments, making it the largest public corpus for sentiment analysis to date. Corpus validation was done using a deep LSTM classifier. This work is done and is yet to be published.

* Principles of Computer & Programming: 19.75/20, Advanced Computer Programming: 20/20, Data Structures & Algorithms: 19/20, Engineering Statistics: 19.5/20, Algorithms Design: 20/20, Principles of Database Design: 18.7/20, Database Design Lab: 20/20, Topics in Computer Science: 20/20, Research & Technical Presentation: 20/20, Artificial Intelligence & Expert Systems: 14/20 (I was sick on the final), Operating Systems:17.4/20, Operating Systems Lab: 20/20, Computer Architecture: 19.6/20, Engineering Ethics: 20/20

- **A Novel Incremental Method For Clustering Temporal Graphs**, Sina Baharlouei, **Ali Janalizadeh Choobbasti**, Maryam Amirhaeri. Jan 2018 – Present
  This work is being done at the Big-data R&D lab at the computer engineering & IT dept. at Amirkabir university of technology. This work is done and is yet to be published.

ONGOING PROJECTS

- **Leader at Data Gathering Group for the Farsi language**, Jan 2018 – Present
  Data is the key to machine learning research, and yet there aren't many large corpora available for deep learning research for the Farsi language. A team of skilled engineers has gathered at Amirkabir University of technology with the goal of scraping and crowd-sourcing the most massive corpora available for Farsi in different research fields. The projects executed by this team are under the supervision of Dr. Reza Safabakhsh, Dr. Saeid Safavi, and Dr. Mehdi Rasti.

  – **FarsChat: A Farsi Text Dialogue Corpus For Research on Multi-Turn Dialogue Systems**, **Ali Janalizadeh Choobbasti**, Armin Kazemi, Reza Safabakhsh. May 2018 – Present
  – **IRVoice: A Multi-lingual Speech Corpus For Research and Development**, **Ali Janalizadeh Choobbasti**, Erfan Gholamian, Mehdi Rasti , Saeid Safavi. May 2018– Present

HONORS & AWARDS

- **Best Paper and Presentation Award**, International Conference on Robotics and Machine Vision (ICRMV). 2017
- **Finalist** in **stock market prediction challenge**, Problem.ir. 2017
- Ranked **29**th amongst the **200** teams participating in the task of revenue prediction, Data Mining Cup (DMC). 2017
- **Ranked Top 4 in term of GPA** among all students in the field of Information Technology, Amirkabir University of Technology. 2014 – Present
- Ranked **340**th among near **250,000** participants in the National Entrance Exam among all Iranian Students in Math. & Physics. 2014
- Ranked **46**th among near **127,000** participants in the National Entrance Exam among all Iranian Students in Foreign languages (English). 2014
- Selected as a qualified person at the first and second stage of **Iranian National Computer Olympiad** 2012

PROFESSIONAL EXPERIENCE

- **Part-time Data Scientist and Big Data R&D Engineer at Miras Technologies International** Jun 2017 – Present
  – I brainstorm along with other lab partners on data-related problems at the **Miras R&D lab**.
  – single-handedly researched, developed and presented the proof of concept for Miras technologies **dynamic pricing** solution. Details cannot be disclosed due to company NDA.
  – R&D on **video advertisement recognition** module – developed a video fingerprinting module.
  – R&D on **document sentiment analysis** module – developed multiple deep learning models (e.g., convolution based, LSTM based, and combinations of the two).
  – R&D on **document deduplication** module – Implemented a document fingerprinting module with a special hash function.
  – R&D on **document topic extraction** module – Implemented an LDA module for topic extraction.
  – wrote multiple **web-scrapers** for data gathering purposes

TEACHING ASSISTANTSHIPS

– **Fundamentals of Data Mining** Dr. Ehsan Nazerfard Feb 2017 – Jul 2017
– **Advanced Computer Programming** Dr. Seyed Majid Noorhosseini Feb 2017 – Jul 2017

– **Computer Architecture** Dr. Saeid Shiri Ghidary — Feb 2017 – Jul 2017
– **Technical English** Dr. Saeedeh Momtazi — Feb 2017 – Jul 2017
– **System Analysis and Design** Mr. Bahman Pourvatan — Feb 2016 – Jul 2016
– **Principals of Programming** Dr. Saeid Shiri Ghidary — Feb 2016 – Jul 2016
– **Data Structures & Algorithms** Dr. Mehdi Dehghan Takht Fooladi — Feb 2016 – Jul 2016
– **Advanced Computer Programming** Mr. Bahman Pourvatan — Feb 2016 – Jul 2016
– **Principals of Programming** Mr. Bahman Pourvatan — Feb 2015 – Jul 2015

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C/C++ , Scala, MATLAB
- **Frameworks and Tools:** Tensorflow, Theano, Keras, Scrapy, NLTK, Scipy, OpenCV, Git, Akka, Docker, OpenGL, LaTeX
- **Database Systems:** Elasticsearch, MySQL

SELECTED ACADEMIC PROJECTS

- **Final Thesis**
  - Design and Implementation of an **attention-based SEQ2SEQ LSTM dialogue system**
    - This project is in progress. I am currently scraping as much dialogue data as I can from Farsi movie subtitles. The NMT-based dialogue system has already been implemented and tested on a small corpus of **200,000** utterances scraped from my personal Telegram messenger chat data where a BLEU score of **3.2** has been achieved.
    - Under supervision of Reza Safabakhsh
- **Kaggle Competition**
  - Convolutional Neural Network for Human Facial Expression of Emotion Recognition.
- **Fundamentals of Data Mining(Audited)**
  - Developed a model for predicting which landing page works best for an E-Commerce website, applied an Artificial Neural Network model for K-class classification using a Softmax layer. Model validation was then done using Bayesian A/B testing.
- **Personal Projects**
  - Design and implementation of a chess playing agent. The agent used a Min-Max algorithm with alpha-beta pruning. The engine has over gone perft testing to make sure of its soundness. Board representation was done using the UCI protocol.
  - Design and development of an Ethical Key-logger (Code not made available for obvious reasons).
  - Design and development of a Super Mario brothers clone. Implemented with the use of the LibGDX game engine.
- **Artificial Intelligence & Expert Systems**
  - Design and development of a search problem solving API.
  - Design and development of an agent in UC Berkeley's famous Pacman competition (Code was not made public as per request of professor). This agent used Monte-carlo tree search and won every other agent designed by fellow classmates.
- **Data Structures & Algorithms**
  - Design and development of an API for graphs and adding an option for path-finding, search and returning its clustering coefficient, solving max flow min cut, topological sort, and the minimum spanning tree of a valid graph.
- **Advanced Computer Programming**
  - Design and development of an Online Strategic Game. A multiplayer online strategy game that also had single player mode played against one or more agents.
  - Design and development of a graphical game named JRobokill with semi-intelligent agents.

REFERENCES

Available Upon Request.