JEREMY TRAN linkedin.com/in/jtran0321
Costa Mesa, CA • 714-800-9638 • jtran0321@gmail.com github.com/jt0321

SUMMARY

Data engineer with extensive knowledge of the entire data science project life cycle, from constructing ETL pipelines to scaling machine learning models in production. Diverse skillset from a background in various fields, such as healthcare and customer management. Coding in Python since 2005.

SKILLS

Expertise in (but not limited to)

- **Programming Languages:** Python, Java, R
- **Deep Learning Frameworks:** PyTorch, MXNet, TensorFlow (Keras)
- **Statistics and Machine Learning Techniques:** linear and logistic regression, A/B testing, gradient boosting
- **Databases:** SQL (PostgreSQL, MySQL, MS SQL), NoSQL (Cassandra, MongoDB, Redis)
- **Big Data and Distributed Computing:** Apache Spark, Hadoop, Airflow, Kafka
- **Visualization:** Jupyter, Matplotlib, ggplot2, Tableau, Excel, PowerBI

Knowledge of

- **DevOps:** Git, Docker, Kubernetes
- **Cloud Computing Services:** Amazon Web Services, Google Cloud Platform
- **Other Programming Languages:** C, C++, JavaScript, Scala

PROJECTS

Jigsaw Unintended Bias in Toxicity Classification (May – June 2019)

- Employed the latest developments in NLP (natural language processing) research to fine-tune neural network models to identify toxic statements with 92% accuracy.
- Methods included transfer learning with BERT (Bidirectional Encoder Representations from Transformers) in the MXNet framework.

Histopathologic Cancer Detection (Jan 2019 – March 2019)

- Trained a convolutional neural network in PyTorch to identify metastatic tissue with 97% accuracy.
- Ranked within the top 7% on Kaggle leaderboard.

EXPERIENCE

Veterans Affairs Medical Center - *Inpatient Pharmacy Technician* Long Beach, CA (Jan 2010 – Apr 2018)

- Assisted in crucial inpatient and inventory operations, including finding cost-saving measures and analyzing expenditure trends for a department with a $24 million annual budget.
- Contributed to workflow redesign that resulted in 40% reduction in the occurrences of missed or delayed scheduled medication administration.

Dr. Tahir Andrabi - *Programming Assistant* Claremont, CA (Jun 2005 – Oct 2005)

- Developed a GUI frontend in Python for data mining large files by adapting implementations of string metrics from the FEBRL project (Freely extensible biomedical record linkage).

EDUCATION

Udacity - *Data Engineering Nanodegree* Online (Apr 2019 – Aug 2019)
Oregon Health and Science University - *Clinical Informatics* Portland, OR (Sep 2011 – Mar 2013)
Pomona College - *Neuroscience, BA* Claremont, CA (Aug 2003 – Sep 2007)