Harrison S. Jansma

https://harrisonjansma.com
(210) 251-1512 | harrisonjansma@yahoo.com

EDUCATION

The University of Texas at Dallas (Part-Time) **Aug 2018 – May 2020**
MS Computer Science, Data Science Track

Baylor University **Aug 2013 – May 2017**
BBA Business Fellows, Mathematics

TECHNICAL SKILLS

- Highly proficient (2-3 years): Python, TensorFlow, Scikit-Learn, MySQL, Pandas, Linux Bash, Git, AWS, Jupyter Notebook
- Proficient (<2 year): Java, Scala, HIVE, Spark, Docker, Redis, JIRA

WORK EXPERIENCE

Data Science Intern – Sprint Overland Park, Kansas **May 2019 – Nov 2019**
Instrumental to the data collection, design, and implementation of a machine learning application which proactively predicts logic failures in Sprint's billing systems.

- Created high performance HIVE and SQL ETL pipelines to aggregate data from various departments. Used aggregated data in design method of novel ML solutions and analytics.
- Prototyped a tree-based model to predict billing errors (Using Scikit-Learn/XGBoost). New model was able to detect four times as many issue types as current models in production.
- Worked with development teams to deploy model to an AWS and Spark production environment. Utilized Linux servers, SSH, and Git to develop production ready code.

Machine Learning Consultant - Upwork Plano, Texas **Oct 2018 – Aug 2019**
Long-term NLP research project with Sociologists at the University of Pennsylvania. Designed a ML model to identify the emotional content of social media profiles.

- Designed Python scripts to collect data from millions of social media profiles. Scripts were faster and able to collect data further back in time than the official Twitter API.
- Created large scale data cleaning and processing pipelines for gigabytes of text data.
- Implemented an emotion-classification model with Gensim, NLTK, and TensorFlow/Keras. Performed statistical analysis of model predictions to show a significant relation between political scandals and population-wide emotional shifts.

PERSONAL PROJECTS

- Scraped 1.2 million articles from Medium.com. Used Docker and AWS to parallelize data extraction and processing.
- RNNs, LSTMs, and Attention Mechanisms for Language Modelling (PyTorch) - Tested the use of word embeddings with a variety of recurrent neural networks towards the task of language modeling (predicting the next word in a sentence).
- Advanced research in neural network architecture for NLP and computer vision. Implemented dozens of research advances in base TensorFlow.