Harrison S. Jansma

https://harrisonjansma.com

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

Data Scientist • NLP Specialist

EDUCATION

The University of Texas at Dallas   **Aug 2018 – Dec 2020**
Master's in Computer Science, Data Science Track

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

TECHNICAL SKILLS

- Highly proficient: Python, NLP, Machine Learning, SQL, Statistical Methods
- Proficient: C++, Versioning (Git), Linux, Spark

WORK EXPERIENCE

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

- Designed and taught a high-level class on machine learning for project managers. Used this time to further discuss their goals for a full-fledged machine learning application.
- Wrote and optimized complex SQL queries to aggregate billing data from multiple tables.
- Prototyped a tree-based model (XGBoost) that predicted customer billing errors with high precision.
- Collaborated with development team to deploy on remote servers using HIVE and PySpark.

Data Science Consultant - Upwork   Plano, Texas   **Oct 2018 – Aug 2019**
Long-term NLP research project with Sociologists at the University of Pennsylvania. Implemented a machine learning model to identify the emotional content of social media profiles.

- Designed Python scripts to scrape textual data from millions of social media profiles.
- Implemented an emotion-classification model based on cutting-edge NLP and deep learning methods.
- Ran predictions on data, then analyzed the results to measure population-wide changes in emotional states after traumatic political events.

Instructional Associate – General Assembly   Plano, Texas   **Mar 2019 – June 2019**
Taught data science fundamentals and best practices to data analytics professionals at Intuit.

- Taught a two-hour class focusing on the application of logistic regression and classification metrics with Python. Audience was a class of thirty senior data analysts.

PERSONAL PROJECTS

- Scraped 1.2 million articles from Medium.com. Created visuals and performance metrics.
- Implemented global average pooling, dropout, and batch normalization in TensorFlow.
- Performed topic modelling with Latent Dirichlet Allocation on comments from Wikipedia.com.
- Implemented K-Means clustering on survey responses dealing with company satisfaction.