

Chaitu Konjeti

ckonjeti1@gmail.com ~ (615) 752-0690 ~ <https://chaitukonjeti.me>

WORK EXPERIENCE

General Motors, Austin, TX

July 2021 - Present

Data Scientist

- Led the pilot development of a conversational chatbot using Langchain and GPT-4 to automate the generation and execution of SQL queries based on questions from the Portfolio Planning division to improve workforce efficiency by 45%
- Developed a survival analysis model to predict the scrappage rate for all vehicles on an annual basis with 73% accuracy in order to provide business intelligence into the demand for non-collision part repairs
- Led the development of a recommendation engine using statistical methods to boost the recommendation rate for vehicles that were previously constrained for dealerships
 - Nationally-launched for all GM dealerships in USA
- Developed a random forest model with 63% accuracy to determine the amount of delay that a vehicle will experience during the delivery to the customer

Social Dynamics and Wellbeing Lab, Georgia Institute of Technology

January 2019 – December 2020

Research Assistant

- Used NLP techniques to train supervised machine learning models on glassdoor posts to predict company performance
- Processed and labeled social media posts to understand people's expression of major life events through social media
- Scraped Twitter, Reddit, and Tumblr posts that contain CDC-provided keywords that correlate with suicide

Operations, Information, and Decisions Department, University of Pennsylvania

May 2019 – September 2019

Research Assistant

- Analyzed the impact of airballs on a basketball players performance using SportVU camera data from NBA games
- Developed software to identify airballs using player and ball coordinates from SportVU camera data
- Correlated changes in performance measures, such as shots, rebounds, passes, with airballs

PROJECTS

Twitter Data Pipeline for Barack Obama's Tweets using Apache Airflow

July 2024

- Ongoing project to extract and transform Barack Obama's tweet data using Tweepy API and Python
- Deployed code using Airflow/EC2

End-to-End ELT Pipeline Using dbt, Snowflake, and Apache Airflow for Customer Sales Data

June 2024

- Configured Snowflake environment with warehouse, roles, schemas, databases, and users
- Established data models, macros, and generic and singular tests within dbt and orchestrated dbt code with Airflow

Song Lyrics Generation using LSTM Network

May 2020 - July 2020

- Ongoing project to develop an LSTM network that takes an input of song lyrics and outputs a randomly generated set of new song lyrics

PUBLICATIONS

Development of a Machine Learning Model Using Multiple, Heterogeneous Data Sources to Estimate Weekly US Suicide Fatalities

Dec 2020

- Journal Publication: [JAMA Network](#)

Crime Prediction and Optimal Police Allocation

May 2018

- Journal Publication: [Young Scientist Journal](#)

Development of basketball metrics using SportVU, an advanced optical recognition system

May 2016

- Tennessee Junior Academy of Science Symposium: Runners Up Award
- Joint Retreat Research Symposium

SKILLS

Data Engineering: dbt, Snowflake, Apache Airflow, SQL (MySQL, Postgres), Spark, ETL Pipelines

Data Science: Python (Scikit-Learn, PyTorch, Pandas, NumPy, PySpark), Langchain, Git

EDUCATION

Georgia Institute of Technology, Atlanta, GA

June 2018 – May 2021

- Bachelor of Science in Computer Science
- Specialization in Artificial Intelligence and Human-Computer Interaction

U.S. Citizen