

# Sanjeev Penupala

@ sanxpen1010@gmail.com

(469) 713-7399

Dallas, TX

in linkedin.com/in/sanjeev-penupala

github.com/spenpal

## EDUCATION

M.S. in Computer Science  
The University of Texas at Dallas

Expected Dec 2023

- Data Science specialization

## SKILLS

### Languages

Python MySQL Bash/Shell

JavaScript Java C++ C

MATLAB R

### Frameworks/Libraries

pandas NumPy Matplotlib

Beautiful Soup scikit-learn Flask

NLTK spaCy

### Technologies

Git UNIX REST APIs

## ACCOMPLISHMENTS

- ACM Projects 2019: **1st Place**
- HackUTD Hackathon 2019: **2nd Place**
- T-Mobile Hackathon 2018: **Top 10 Finalist**

## COURSEWORK

Artificial Intelligence

Database Design

Data Structures and Advanced Algorithms

Design & Analysis of Algorithms

Machine Learning

Natural Language Processing

Advanced Operating Systems

Probability and Statistics

Software Engineering

## WORK EXPERIENCE

EDG Intern

Mathworks

September 2021 - December 2021

- Tools - MATLAB, JavaScript
- Developed a design space preview for simulations in Simulink, enabling users to visualize parameters in the design space, decreasing parameter testing time.
- Created over 10 unit tests to ensure optimal performance of the codebase.
- Integrated parameter grouping into the Simulink UI, enabling users to build simulation runs for a vast array of scenarios efficiently.

Software Engineer Intern

Walmart Global Tech

June 2021 - August 2021

- Tools - Python, Facebook Prophet, Azure Databricks
- Crafted an end-to-end forecasting system for estimating future cloud computing costs, resulting in a 5% reduction in annual expenditure.
- Designed 6 unique metrics for assessing cloud computing costs and identifying areas of overspending.
- Generated an interactive Databricks dashboard for the evaluation of multiple forecasting models, achieving an accuracy from 69% to 81%.

Undergraduate Researcher

NSF REU at the University of North Texas

May 2020 - August 2020

- Tools - Python, Matplotlib, pandas, scikit-learn, Doccano
- Engineered a 3-layered classification pipeline, boosting the accuracy of 7 classifiers from 80% to 95% for identifying argumentation in legal corpora.
- Oversaw dataset quality control by remodeling annotation guidelines, increasing annotator agreement by over 40%.
- Optimized data cleaning scripts to remove 700+ conflict labels for re-labeling.

## PROJECTS

Volleybot

Chatbot

- Tools - Python, Beautiful Soup, NLTK, spaCy
- A simple chatbot that answers any questions related to volleyball!
- Performed web crawling & scraping from over 100+ sites to create a detailed knowledge base, using TF-IDF scores, of over 500+ facts and statements.
- Utilized vector cosine similarity and POS tagging to retrieve the most relevant response, from the knowledge base, to a lemmatized user query.

HashPost

Social Media App

- Tools - Python, Flask, Flutter, Firebase, Imagga API, Hashtagify API
- Mobile app that generates hashtags and captions for social media posts.
- Developed algorithm to generate unique URLs for images uploaded by users.
- Created tagging system, generating trendy hashtags from objects in the image.