

# PANKAJ MANOHARLAL THAKUR

+1 9195597560 | [pmanoha@ncsu.edu](mailto:pmanoha@ncsu.edu) | [linkedin.com/pankajthakur62](https://www.linkedin.com/pankajthakur62) | [github.com/PankyThakur](https://github.com/PankyThakur)

## Education

North Carolina State University, Raleigh, NC

Aug 2023 - May 2025

*Master of Computer Science*

GPA: 4.0/4.0

Courses: Automated Learning and Data Analysis, Generative AI for Software Engineering, Internet Protocols, Neural Network and Deep Learning, Foundations of Data Science, User Experience

Siddaganga Institute of Technology, Tumakuru, India

Aug 2016 - Aug 2020

*Bachelor of Engineering in Electronics and Communication*

GPA: 9.2/10.0

Courses: Data Structures, Data Analytics, Object Oriented Programming, Machine Learning, Advanced Image Processing

## Work Experience

Generative AI Research Assistant under Dr.Bowen Xu

Sep 2023 - Present

*Software Engineering Lab, NC State, Raleigh, NC*

- Conducting **Sentiment and Data Analysis** on reviews, fine-tuning **Large Language Models** including **GPT-4** to cluster the reviews. This process is aimed at examining developer sentiments related to Generative AI-based extensions.
- Enhanced an existing data crawler to automate the collection of extension reviews, resulting in a **70%** time savings. Additionally, created **Taxonomy** of the user reviews.

Network Consulting and Automation Engineer II

Jan 2020 - Jul 2023

*Cisco Systems India Pvt. Ltd, Bengaluru, India*

- Designed and developed the frontend and backend of a tool that reduces the time to migrate the Cisco Contact Centre data, from **2 months to 72 hours** using Python, Django, and AXL APIs.
- Devised scripts for automating network data collection across **485 devices** making use of Python and Jupyter Notebook, achieving a **40%** reduction in time. Received **connected recognition** from a senior architect for the successful implementation.
- Automated the backup process for more than **10,000** UC devices using Python, SQL, and Jupyter Notebooks, leading to a time-saving of **35%**. The scripts were used in Banking and Public Health organization projects.

*Network Consulting and Automation Intern*

- Revamped and engineered a query tool by employing Angular, Java, SpringBoot, REST APIs, SQLite, and a **CI/CD pipeline**. Achieved a **30%** enhancement in user experience, introduced supplementary functionalities, and bolstered scalability.

## Projects

**Natural Language Processing**

- Built classifiers employing frequency-based, LSTM and BERT methods in **Python** with the help of **PyTorch** to classify tweets into 5 labels(Extremely negative, Negative, Neutral, Positive, Extremely Positive) achieving an accuracy of 64%, 73% and 85.75% respectively for the three methods and **published a paper** on the same. [Link](#)

**Data Analytics**

- Constructed dataset with **12** parameters with the help of ETL process, for oil price prediction, by collecting data from diverse sources. Employed rigorous feature selection techniques to choose 6 key features, resulting in a Mean Squared Error of **3.33** for the KNN + Random Forest ensemble model.

**Generative AI**

- Spearheaded "README2Video" project utilizing Large Language Model - **Falcon** and **ChatGPT API** to seamlessly convert README files into engaging video prompts through a novel video pipeline. Successfully generated videos for CSC faculty members using our pipeline.

**Machine Learning**

- Engineered a sophisticated credit risk prediction model in Python with scikit-learn, integrating **CART, NB, and XGBoost** models to build a pipeline. Achieved an impressive F1 score of **74%**, leveraging historical financial data and implementing advanced feature engineering

## Technical Skills

**Programming Languages:** Python, C, C++, Java, R, XML, HTML

**Databases:** SQL, SQLite, MySQL

**ML & DS Skills:** Exploratory Data Analysis, Visualization, Data Transformation, Modeling, Feature Engineering, Recommender Systems, Natural Language Processing, Linear Regression, Neural Networks, SVM, Dimensionality Reduction, Large Language Models, Deep Learning, Computer Vision, Data Modeling, Data Mining, Bayesian Regression Models, NLTK, Spacy,Scikit-Learn, Matplotlib

**Frameworks & Libraries:** Angular, JavaScript, Django, Flask, Numpy, Pandas, REST API, Agile, AWS S3, Lambda, LLMs

**Tools:** Eclipse, Power BI, Excel, Jupyter , Git, Docker, PyTorch, Tensorflow, SpringBoot, RStudio, PyCharm, Apache Spark

**Soft-Skills:** Leadership, Proactive, Teamwork, Communication, Problem-Solving, Presentation Skills, Business Acumen

## Leadership and Recognitions

**Cisco-SIT Industry-University Connect:** Orchestrated a program facilitating the collaboration between Cisco experts and SIT through seminars, projects, and research problem statements. Engaged **more than 500 students**, enhancing their performance in campus placements. Recognized with the **Emerging Talent Award** for significant contributions.

**Vice Chairman, Robotics Club:** Led and mentored a **team of 60** in the construction of state-of-the-art robots. Organized three national-level robotics championships over three years, attracting **700+** students annually.