

# Sumanth Gopalkrishna

3220 E John Hinkle Pl Apt D, Bloomington, Indiana 47408

Ph: 812-369-3116 | [sgopalk@iu.edu](mailto:sgopalk@iu.edu) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

<b>Master of Science, Data Science</b>   Indiana University Bloomington, USA	<b>Dec 2022</b>
<i>Courses: Machine Learning, Artificial Intelligence, ML in Bioinformatics, Deep Learning, Big Data</i>	GPA: 3.9/4.0
<b>Bachelor of Engineering, Computer Science</b>   Bangalore Institute of Technology, India	<b>May 2018</b>
<i>Courses: Databases, Data Structures, Algorithms, Web Development, Data Mining</i>	GPA: 3.8/4.0

## TECHNICAL SKILLS

- **Programming languages:** Python, SQL, R, JavaScript, C, C++
- **Database and Web Technologies:** MySQL, SQLServer, HTML5, CSS, Flask, JSON, Alteryx, UiPath
- **Machine Learning:** Regression, Classification, Clustering, Decision Trees, Random Forest, Ensemble models, PCA, SVM, KNN, Recommender Systems, Neural Networks, CNN, RNN, LSTM, XGBoost, LGBM
- **Libraries:** Tensorflow, Keras, Numpy, Pandas, Scanpy, Scipy, Matplotlib, Scikit-learn, Nltk, AnnData, Ggplot, PyTorch
- **Data Visualization:** Tableau, Matplotlib, Seaborn, Plotly

## WORK EXPERIENCE

**Graduate Research Assistant** **May 2021 – Present**  
*Indiana University | Bloomington, IN, USA*

- Extracted PDF organizational tax forms to tabular format and pre-processed the data using **NLP** techniques to conduct in-depth **network analysis** and analyse the grantee-foundation relationships in the U.S. NPO Sector.
- Built dynamic **network graphs** in **Tableau** to visualize the cash flow and associations between foundation and grantee organizations and maintain transparency to the US citizens.

**Data Scientist** **Aug 2018 – Jan 2021**  
*Moog Inc. | Bangalore, India*

- Analysed aircraft actuator data to predict the actuator failure rate with features such as friction, pressure, and velocity using **XGBoost** with recall of 0.87 and F1 score of 0.81.
- Developed and designed multiple visualization dashboards using **Node.js** which was used by C-level and upper management to make data driven business decisions.
- Developed an automated database backup mechanism using **MySQL** procedures and triggers thereby saving 100 hours of manual work per month.
- Implemented a **Robotic Process Automation (RPA)** process to automate applications using the **UiPath** tool to reduce manual workload by of HR, sales, and medical in-house teams by 60%.

## PROJECTS

### E-Commerce Recommendation System | Recommendation Engine, Collaborative Filtering

- Built a recommendation engine comprising of a **collaborative filtering model** using stochastic gradient descent and alternating least squares optimization algorithms for an E-Commerce website on the user review data.

### Question Answering System | Natural Language Processing, Neural Networks

- Built a **question answering (QA)** system using **Siamese network** and **Bert** to predict the answers for the google search queries with an F1 score of 0.75.

### Credit Card Fraud Detection | Classification Model, XGBoost, Random Forest

- Identified fraudulent credit card transactions with classification models such as **XGBoost**, **LGBM** and **Random Forest**. The XGBoost model achieved the best performance with a recall of 0.80 and AUC score of 0.85.

### Horizon Detection | Viterbi algorithm, Flask

- Deployed a webapp that captures the horizon of any image using Viterbi HMM algorithm and ImageIO on **Heroku** server using **Flask**.

### Part of Speech Tagging | Markov model, Gibbs Sampling

- Implemented **part of speech (POS)** tagger on the document texts using **Bayes Net** and **Gibbs Sampling** with an accuracy of 93.5%.