

Shubhkirti Prasad

+1 (812) 671 5196 | shubpras@iu.edu | <https://shubhkirti24.github.io/> | <https://github.com/Shubhkirti24> | <https://www.linkedin.com/in/shubhkirti-prasad/>

EDUCATION

Indiana University Bloomington, MS in Data Science | Indiana, USA GPA: **3.7 / 4.0** May 2024
Manipal Institute of Technology | Karnataka, India GPA: **8.3 / 10** Aug 2021

Competitions: •BAJA Virtuals India (2019-20) •BAJA Pithampur 2020 •ESI 2020,Pune •BAJA Louisville 2020, USA
Courses: Statistics | Advanced Database Concepts | Applied ML | Computer Vision | Data Visualization | NLP | MLOps

EXPERIENCE

Luddy School of Computing Associate Instructor(Usable Artificial Intelligence) | Indiana University, USA Jan 2024 - Present

- Natural Language Processing, Problem Solving, Machine Learning, A/B testing, Leadership, Mentoring
- Facilitating a comprehensive understanding of machine learning algorithms, to over 100 graduate students.
- Conducting research, designing, analyzing, and visualizing data sets, helping students navigate through a fast-paced timeline.

Paul O'Neill School, Luddy School Graduate Teaching Assistant (V506 :Statistical Analysis) | Indiana University, USA Aug 2023 - Dec 2023

- R, Statistical Modeling, Data distribution, Teaching, Hypothesis testing, A / B testing, MLS
- Facilitated statistical modeling in R for over 40 graduate students, contributing to a class average grade of A.
- Informatively integrated real-world data sets (OpenIntro Statistics) into weekly lab sessions, enhancing practical understanding and application of statistical theories.

Four Colors Technology, Data Science Intern (LLM Modeling) | Remote (MA, USA) May 2023 - July 2023

- NLP, Langchain, LLM, SpaCy, Python, Named Entity recognition, Vertex AI, Google Cloud Platform, Prompt Engineering, RAG, Model Finetuning
- Spearheaded the customization of Google's Bison LLM, to be used as a backbone for employee support chatbot on Vertex AI
- Utilized various techniques like RAG(Retrieval Augmented Generation) using Prompt Engineering, Document chunking, Vector Space Embeddings effectively on Vertex AI to enhance accurate response generation.
- Developed a context-based model that enhanced the accuracy of an chat-bot by over 30%. during A/B testing.

DataOrc, Data Science Intern (OCR / SQL) | Remote (Pune, India) AUG 2021 - NOV 2021

- Python, Convolutional NN, Jupyter Notebook, Sklearn, NumPy, Pandas, TensorFlow, Keras
- Played a key role in a team that developed an OCR system, translating user's handwritten domain-specific text into English with 72% accuracy using CNN.
- Optimized SQL queries with Relational Algebra (decreasing the overall query runtimes by 32% to aid the dry run and deployment of a highly scalable, data ready analytics platform

IBM, Data Science Intern (Financial Analysis) | Remote (Bangalore, India) APR 2021 - JUL 2021

- Python, Statistical Analysis, Time Series, Sklearn, NumPy, Pandas, data mining, RegEx, Data Pipelines
- Identified and researched a business problem to evaluate the health of of BSE-500 companies Pre/Post Covid.
- Created a Data pipeline to extract relevant data from Annual Reports of BSE-500 companies for FY-18,19,20,21.
- Lead a project 'Distribution in IT Sector with Investment Impact Analysis', a comparative analysis of firms in manufacturing sector trading in BSE-500 for linkage between IT investment and the firm's financial performance.

SKILLS

Languages Python(TensorFlow, Keras, Pandas, Scikit-learn, etc.), PySpark, R, C/C++, Bash
Vertex AI, Google Cloud Platform, Huggingface, AWS, TensorFlow, Pytorch, Docker, Jupyter Notebook,
Software Anaconda, PostgreSQL, Tableau, Azure SQL Database, Microsoft Office (Excel, Word, Power Point),
Solidworks
Certifications Lean Six Sigma yellow belt (Govt. of India) | Professional Data Scientist, IBM – (2021)
Soft Skills Leadership, Time Management, Problem-solving, Documentation, Presentations

PROJECTS

News Maker's Network | NLP November '24

- Skills : Vector Databases, Transformers, Named Entity Recognition with BERT Transformer, Spacy Transformer, Sentiment Analysis, Network Models, Flask, HTML, NLTK
- Created a novel tool for US elections that utilised a transformer based model to extract named entities from over 9,000 news articles to create a network model of related entities
- The model generates a sentiment graph of a co-mention network
- [Click Here to go to the website](#)

Tropical Storm Intensity Estimation | CNN

April 23

- Created a novel CNN model (Classifier + Regressor) based on VGG-16 to estimate the wind speed of 600 individually documented storms with their cloud formation progression (100k) images(RAW + Infrared Images)
- Optimized the model using Data Augmentation (Image Generator), Max Pooling, Dropouts, Batch-Normalization, Convolutional layers, activation functions, fully connected layers, image filter, edge case hypothesis.
- [Click Here](#) to go to the paper

Job-Skill Dashboard | ONET, Tableau

April 23

- Created an interactive job skill mapping dashboard using Tableau based on the ONET database to visualize the current employment characteristics in the USA, based on parameters such as Age, Location, Skills, Knowledge, Education, Experience, etc.
- Implemented advanced SQL optimization techniques (Relational Algebra, set operations and SQL triggers) to reduce query executing time from 12 to 2 Mins (-83%) on a dataset containing 1000+ occupation categories over 1,000,000+ rows.
- [Click Here](#) to go to the project github

Hand Drawn Image Classification | CNN

November 22

- From a hand drawn dataset of 500,000 images with 100 classes, classified 100,000 images using a custom CNN model with an accuracy of 72%
- Scored 2nd out of 100 models, Applied techniques: Data Augmentation, Max Pooling, Dropouts, Batch-Normalization

Mechanical Projects - Team Manipal Racing

Jan 2018 - April 2021

- Research, Development and Manufacturing of an off-road car (BAJA buggy) - Research, manufactured and optimized custom components of a CVT transmission system and a custom transmission assembly.