



Shubhkirti Prasad

+1(812)6715196 | shubhkirti.prasad@gmail.com |  shubhkirti24.github.io/ |  linkedin.com/in/shubhkirti-prasad/

EDUCATION

Indiana University Bloomington, MS, Data Science Indiana, USA	GPA: 3.7 / 4.0	May 2024
Stanford University, Machine Learning	Grade: A+	Mar 2024
Manipal Institute of Technology, BTech, Mechanical Karnataka, India	GPA: 8.3 / 10.0	Aug 2021





EXPERIENCE

- Indiana University, Associate Instructor (Usable AI, Statistical Analysis)** Aug 2023 - May 2024
- Leveraged expertise in Natural Language Processing and Machine Learning to mentor over 100 graduate students in A/B testing methodologies and problem-solving; led research initiatives and developed Gen AI learning modules on Prompt Engineering, LangChain, Huggingface, and Model Finetuning.
 - Taught statistical modeling using R, data modeling, and hypothesis testing; integrated real-world datasets into the curriculum to enhance students' proficiency in MLS and A/B testing.
- Four Colors Technology, Data Science Intern (LLM Modeling)** | MA, USA May 2023 - Jul 2023
- Led the adaptation and finetuning of Google's PaLM 2 on Vertex AI for a support chatbot, utilizing Langchain and model finetuning strategies, achieving over 30% improvement in response accuracy through A/B testing with RAG, Chain of Thought, Zero-Shot, and ReAct prompt engineering strategies.
- DataOrc, Data Science Intern (OCR / SQL)** | Pune, India Aug 2021 - Nov 2021
- Developed an OCR system using a Convolutional Neural Network that achieved 72% accuracy in translating domain-specific handwritten text into English and enhanced SQL query performance by 32% through Relational Algebra, supporting the deployment of a scalable, data-ready analytics platform.
- IBM, Data Science Intern (Financial Analysis)** | Bangalore, India Apr 2021 - Jul 2021
- Conducted an in-depth pre/post-Covid data analysis of BSE-500 companies, revealing a 12% decrease in overall performance using ETL, Ratio Analysis, Trend Analysis, ANOVA, and Discounted Cash Flow Analysis; also led a project analyzing the correlation between IT investment and financial performance in the manufacturing sector over four fiscal years, creating data pipelines for data mining and statistical analysis, which improved efficiency by 38% for over 50 listed companies.

SKILLS

Programming Languages	Python(TensorFlow, Keras, Pandas, Scikit-learn, etc.), SQL, R, C/C++
Databases	PostgreSQL, MySQL, NoSQL(MongoDB, Redis), Azure SQL Database, Oracle SQL Server
Analysis, Visualization	Matplotlib, Seaborn, Tableau, Power BI
Cloud / Software	Vertex AI, Google Cloud Platform, Huggingface, AWS, Docker, Conda, Solidworks, Git
Certifications	Lean Six Sigma yellow belt (Govt. of India) Professional Data Scientist, IBM - (2021)
Soft Skills	Leadership, Collaboration, Time Management, Problem-solving, Documentation
Management Skills	Agile, Lean Six Sigma

PROJECTS

- News Maker's Network | NLP | Website**  Nov 24
- Skills : Vector Databases, HTML, NLTK, Transformers, Named Entity Recognition with BERT Transformer, Spacy Transformer, Sentiment Analysis, Network Models, Flask, Docker, Kubernetes
 - Developed of a transformative NLP tool for analyzing US election trends, extracting key entities from 9,000 news articles using a transformer-based model to generate a comprehensive network model and sentiment graph, significantly enhancing data-driven election insights.
 - Deployed on Google Cloud using Docker and Kubernetes, leveraging the scalability and management features of Kubernetes while ensuring that the application was containerized for portability and consistency.
- Tropical Storm Intensity Estimation | CNN | Paper**  Apr 23
- Created a novel CNN model (Classifier and Regressor) based on VGG-16 to estimate the wind speed of 600 individually documented storms with their cloud formation progression (100k) images(RAW and Infrared Images)
 - Optimized using Data Augmentation (Image Generator), handled class imbalance, Max Pooling, Dropouts, Batch-Normalization, image filters, edge case hypothesis.
- Job-Skill Dashboard | ONET, Tableau | Project github**  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.
- Hand Drawn Image Classification | CNN** Nov 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%, scoring 2nd out of 100 models
- Mechanical Projects - Team Manipal Racing | Project github**  Jan 2018 - Apr 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.