ABHISHEK MANOJ SUTARIA

abhishek.sutaria@gmail.com | +1 (930) 333-7194 | LinkedIn: Abhishek Sutaria

EDUCATION

Indiana University Bloomington **Aug 2024 - May 2026**
Master of Science in Data Science Indiana, USA
Relevant Coursework: Data Mining, Advanced Database Technologies, Big Data Management Systems

University of Mumbai **Jun 2018 - May 2022**
Secured an Honors Specialization in **Artificial Intelligence and Machine Learning** CGPA 9.78/10
B.E. in Electronics and Telecommunication Engineering
Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Applied Statistics

TECHNICAL SKILLS

- **Languages**: Python, R, Spark (PySpark), SQL, Java, Shell Scripting, Scala
- **Libraries**: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Polars, Selenium, NLTK, Keras, TensorFlow
- **Data Science**: Machine Learning, Deep Learning, Natural Language Processing, Data Analytics, Business Intelligence
- **Technologies**: Git, Subversion, MySQL, PostgreSQL, Snowflake, Power BI, Tableau, Hadoop, Pytorch, Excel
- **Cloud Computing**: AWS, GCP, Azure, EC2

PROFESSIONAL EXPERIENCE

Edelweiss Global Markets, Mumbai, India | *Data Engineer* **Jul 2022 – Aug 2024**

- Awarded for improving the quality of data by 70%, resulting in a 40% reduction in data-related errors.
- Engineered 40+ ETL data pipeline processes using Apache Airflow, reducing manual processing time by 80% and increasing daily data throughput from 500GB to 1TB.
- Architected a robust Python validation framework using Pandas and Great Expectations, elevating data integrity from 85% to 98% validating 500K+ live market daily data points being stored in Clickhouse.
- Containerized PySpark workflows using Docker and orchestrated them on an 8-node Kubernetes cluster, reducing 500GB data processing time from 5 hours to 2 daily.
- Orchestrated integration of 4+ data sources and also a third-party API using Python and R, streamlining data flow and reducing data latency by 40%.
- Developed a bash script on Linux to automate purging of 10-12TB high-volume tick-by-tick data post-validation against master data stored in AWS buckets improving storage efficiency by 30% and saving 8+ human hours monthly.

Bitgenie Private Limited, Mumbai, India | *Data Science Intern* **Jun 2021 – Jul 2021**

- Performed web scraping of over 50+ websites, extracting 600+ reports by using libraries like BeautifulSoup. Utilized Google Dialogflow to implement a virtual ChatBot.
- Launched **Amazon EC2 Instance** to streamline web-scraping processes, reducing processing time by 30%.

Being Digital, Mumbai, India | *AI/ML Intern* **Mar 2021 – May 2021**

- Trained a Deep Learning model to recognize over 20 classes of vegetables with an accuracy of 94%.
- Deployed a web application on Heroku that scans vegetable images, displays nutritional information, and recommends the top three recipes using HTML5 and CSS3 for the front end and Flask for the back end.

Phemesoftware Private Limited, Mumbai, India | *Python Developer Intern* **Jun 2021 – Jul 2021**

- Architected a high-performance Virtual Workspace using HTML/CSS frontend with Django/MongoDB backend, enabling secure remote collaboration for 500+ freelancers/contractors with 40% efficiency gain

ACADEMIC PROJECTS AND RESEARCH PAPERS

Diabetic Retinopathy Detection using Convolution Neural Networks and DR-GANs **Jun 2021**

- Devised an automated model using CNN and GAN to diagnose and classify diabetic retinopathy from retinal images.
- Utilized Git version control for collaborative model development, achieving 92.3% accuracy in detecting multiple stages of DR severity, surpassing previous automated methods by 15% and processed up to 1000 retinal images per hour, a 10x improvement over manual screening capacity.
- Reduced diagnosis time by up to 60% relative to manual screening methods, enabling timely treatment and reducing vision loss risks.
- Won the IETE-SF's DJ Spark project-based technical paper presentation competition and published in the State-Level Journal DJ Spark.

Elements of Nature Optimized into Smart Energy Grids using Machine Learning **Jan 2022**

- Published a paper on Design Engineering in the international journal in Toronto, June 2021, Issue 5, ISSN: 0011-9342.
- Trained a machine learning model to predict solar radiation for smart energy grids, achieving a 72% cross-validation score using a Hyper-Parametrized Random Forest algorithm, significantly outperforming traditional methods and employed data visualization techniques (confusion matrices, ROC curves) to interpret model performance.
- Implemented and evaluated multiple machine learning algorithms (Linear Regression, Lasso Regression, Random Forest, SVM) improving forecasting accuracy by over 20% compared to baseline models.
- Conducted comprehensive data analysis and feature engineering on weather station data, identifying temperature as the most influential factor for solar radiation prediction.