# Shubbham Gupta

Dublin, Ireland | +353899586207 | shubbham.gupta28@gmail.com | www.linkedin.com/in/shubbhamgupta | Stamp 1G

## OBJECTIVE

PhD-researcher in SFI CRT Data Sciene with extensive deep hands-on experience in predictive analytics, scalable machine learning pipelines, and Bayesian modeling. Proven ability to implement cloud-native solutions in healthcare and pharmaceuticals, apply SOLID principles in software engineering, and build intuitive dashboards for real-time data exploration. Passionate about developing robust AI systems with MLOps/LLMOps frameworks, integrating LLMs, and deploying end-to-end ML workflows in production.

## SKILLS

- **Programming & Databases:** Python (OOP, SOLID, PEP8), R, C++, SQL, MySQL, PostgreSQL, MongoDB
- **Data Analysis & Visualization:** Power BI, Tableau, Plotly, Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost
- **Cloud, Big Data & Development Tools:** AWS, Azure, GCP, Hadoop, Spark, Hive, Jupyter Notebook, RStudio, Anaconda, GitHub, Git, MS Office
- **Machine Learning & AI:** Machine Learning, Bayesian Statistics, Neural Networks, NLP, Artificial Intelligence, Model Training, Hyperparameter Tuning
- **MLOps & LLMOps:** Streamlit, Docker, GitHub Actions (CI/CD), Hugging Face Hub, OpenAI API, FinBERT, LLaMA, Prompt Engineering, Config Management (YAML), Modular ML Systems, Env Variables

## RESEARCH EXPERIENCE

*PhD researcher in Data Science, University College Dublin* — *Dublin, Ireland*
MetaboVariation: Exploring individual variation in metabolite levels — *Sep 2020 - Apr 2025*

- Designed and developed "MetaboVariation" a **novel multivariate Bayesian generalised linear model** to capture intra-individual variations in metabolite levels across repeated measurements.
- Applied **SOLID principles** to develop a **user-friendly R package** with an interactive **Shiny dashboard**, streamlining analysis and enabling clinicians to explore and interpret model outputs easily.
- Collaborated with international researchers (Helmholtz Zentrum München) to integrate real-world use cases.
- Presented insights via interactive dashboards and secured **Best Poster Award** at the 42nd Conference on Applied Statistics in Ireland.
- **Authored peer-reviewed publication** in "Metabolites" and contributed to open-source packages.

*Data Science and Artificial Intelligence research intern, Novartis Ireland Limited* — *Dublin, Ireland*
Analysis of Atherosclerotic cardiovascular disease — *Jun 2021 - Sep 2021*

- Built a **python-based survival modeling pipeline** for analyzing **real-world data** of over **40,000 patients** to identify key predictors of ASCVD outcomes using Cox and parametric survival models.
- Created **interactive dashboards** using Plotly and SHAP for internal review, enabling dynamic filtering by risk factors.
- Followed **SOLID principles** and **clean coding** to develop code integrated into Novartis' internal framework.
- Delivered the application findings of data-driven methods in clinical approaches in a Novartis in-house seminar.

*Research Fellow (Short-term), University College Dublin* — *Dublin, Ireland*
Sheepdog Algorithm — *Jan 2021 - Jun 2021*

- Developed the Sheepdog Algorithm, **a bio-inspired metaheuristic feature selection tool** inspired by sheepdog herding behavior that **reduces dimensionality** by balancing low feature count and high accuracy.
- Integrated the algorithm with **multiple machine learning models** (Random Forest, SVM, Logistic Regression) in Python for dynamic feature pruning, achieving a 10% **improvement in predictive performance** through rigorous benchmarking and A/B testing on real-world biomedical datasets.
- Presented the innovative Sheepdog Algorithm to over 200 professionals at major European and international conferences, showcasing its **novel research findings**.

*Master's researcher, University College Dublin* *Dublin, Ireland*
Survival Analysis of clinical mastitis *April 2020 - Aug 2020*

- Identified key risk factors for mastitis by applying **Cox regression models, mixed-effects models, and L1 penalization**, revealing the most influential covariates affecting the hazard of infection.
- Enhanced early detection strategies for mastitis by **estimating survival curves** and evaluating model fit through **residual analysis**, thereby contributing to better disease management protocols.

## EDUCATION

*University College Dublin* Dublin, Ireland
PhD in Data Science *Sep 2020 - Apr 2025*

- Specialized in predictive modeling, scalable ML, and feature selection techniques. Engaged in advanced research, including the development of novel predictive models for metabolic disorders and innovative feature selection techniques. Contributed to interdisciplinary teaching and collaborated internationally, leading to a publication and award.

*University College Dublin* Dublin, Ireland
MSc Data and computational science *Sep 2019 - Aug 2020*

- Conducted research on survival analysis in healthcare, alongside comprehensive coursework in predictive analytics, Bayesian analysis, and machine learning. Acquired proficiency in multiple programming languages and statistical tools, including R, Mathematica, and Fortran.

*Guru Gobind Singh Indraprastha University* New Delhi, India
B.Tech, Computer Science and Engineering *Sep 2015 - Aug 2019*

- Gained a solid foundation in computer engineering, covering operating systems, web and android development, big data, and programming languages like C++, Java, and Python. Developed skills in database management and software development.

## RELEVANT PROJECTS

- *AI-Powered Financial Sentiment Agent | https://github.com/shubbham28/financial-agent*
  - Developed an **real-time AI agent** integrating **Yahoo Finance, Finviz news** for sentiment analysis, with trend predictions and LLM-powered summaries.
  - Developed chatbot UI using **Streamlit**, enabling user-friendly interaction for multi-stock queries and trend insights.
  - Leveraged **technical indicators** (EMA, RSI, MACD, Bollinger Bands) for **real-time stock analysis**.
  - Applied **FinBERT** for news sentiment extraction and LLaMA + OpenAI for prompt validation and summary generation.
  - Followed **MLOps best practices**: modular architecture, YAML-based config management, environment isolation, and auto-key loading for secure deployment.
  - Implemented **LLMOps workflows**: prompt chaining, validation, and role-based prompting using Hugging Face models and OpenAI APIs.

## PUBLICATIONS

- *MetaboVariation: Exploring Individual Variation in Metabolite Levels*
- *Quantum Grey Wolf Optimization and Evolutionary Algorithms for Diagnosis of Alzheimer's Disease*
- *Bio-Inspired Algorithms for Diagnosis of Breast Cancer*
- *Modified Bio-Inspired Algorithms for Diagnosis of Breast Cancer Using Aggregation*

## INTERESTS

Large Language Models, Cloud Computing, Finances, Stocks, Healthcare, Cycling, Hiking, Running, Badminton