Dr. Yogesh Bansal

Full-Stack ML Engineer | Data Scientist | Innovated Multi-Modal Data Solutions
yogesh.bansal@ucdconnect.ie | linkedin.com/in/dr-yogeshbansal | dryogibansal.github.io

PROFILE

Full-stack ML Engineer and Data Scientist with expertise in designing and developing novel ML model architectures. Experienced in building end-to-end scalable ML pipelines, cloud-based model deployment, and MLOps workflows. Skilled in frameworks like TensorFlow leveraging multi-modal data for predictive analytics and time-series forecasting.

SKILLS

ML Frameworks	TensorFlow, PyTorch, Scikit-learn
MLOps Tools	AWS SageMaker, Docker, Kubernetes, Airflow, CI/CD (GitHub Actions)
Cloud Platforms	Amazon Web Services
Programming	Python
Deployment	API Development (Flask)
Database	MySQL, PostgreSQL
Visualisation, Analytics	Pandas, Matplotlib, Seaborn, Tableau

EDUCATIONAL ACHIEVEMENTS

Course of Study	Educational Institution	Qualification Level (NFQ)	Grade	From	To
Doctor of Philosphy (Computer Science)	University College Dublin, Ireland	Level 10	Passed	01/09/2018	08/05/2024
Master of Science (Computing)	Dublin City University, Ireland	Level 9	2nd Class Honours Grade 1	01/09/2017	31/08/2018
Master of Engineering (Information Technology)	Panjab University, India	n/a	1st Division	01/10/2009	31/07/2011
Bachelor of Technology (Information Technology)	Technological Institute of Textile and Sciences, India	n/a	1st Division with Honours	01/08/2004	31/07/2008

CAREER OVERVIEW / PROFESSIONAL EXPERIENCE

Job Title	Employer	From	To
Data Analyst and Project Lead (Postdoctoral Fellow)	University College Dublin, Ireland	01/11/2023	31/10/2024
Technical Mentor (Demonstrator)	University College Dublin, Ireland	01/09/2018	31/12/2023
Information and Communication Technology Intern	Irish Wheelchair Association, Ireland	11/06/2018	31/08/2018
Big Data Developer, Trainer	Init Call Technologies, India	17/12/2015	31/08/2017
Technical Mentor (Assistant Prof.)	Baddi University, India	02/08/2011	15/12/2015
Web Developer	DesignersX, India	02/03/2009	10/09/2009
PHP Programmer	Megh Softwares, India	08/11/2008	10/02/2009

Detailed Career History and achievements

Job Title: Data Analyst and Project Lead (Postdoctoral Fellow)

- Developed a georeferenced database with over 2M+ records, integrating weather, traffic, and geodirectory datasets to support advanced regional analysis.
- Conducted data preprocessing, visualisation, and analysis to uncover regional business trends and traffic patterns while mentoring junior researchers in data integration workflows.

Job Title: Technical Mentor (Demonstrator)

- Demonstrated practical labs on ML, deep learning, big data, and RDBMS modules to 200+ students annually, aligning sessions with industry-relevant technologies like Python, Hadoop, and MapReduce.
- Mentored students on technical projects, guiding them to successfully implement solutions and complete assignments within deadlines.

Job Title: Information and Communication Technology Intern

- Trained staff in adopting modern ICT tools for better data management, improving their understanding of data consistency and accuracy within the 3-month internship period.

Job Title: Big Data Developer, Trainer

- Optimised distributed data pipelines by reconfiguring MapReduce workflows, reducing query processing times from several hours to under 1 hour for large-scale datasets.

Job Title: Technical Mentor (Lecturer/Assistant Professor)

- Taught programming languages (C, C++, PHP) and data structures to 200+ undergraduate students over four years, designing industry-aligned course modules.
- Successfully mentored 5 postgraduate students on technical thesis projects, all of whom defended successfully.

Job Title: Web Developer / PHP Programmer

- Developed 3+ dynamic websites (both front-end and back-end) for international clients, integrating content management systems like WordPress.
- This enabled clients to easily manage website content, such as adding blog posts, editing pages, and uploading images thereby improving user engagement and content management efficiency.

Projects (Machine Learning)

Project: End-to-End Financial Risk Analysis Pipeline | Dates: 01/11/2024 - Present
Tools and Libraries: Python, PyTorch, Docker, Jenkins, AWS Lambda, Flask

- Currently developing a Financial Risk Analysis Pipeline, leveraging ML to analyse financial data and identify high-risk profiles.

Project: Georeferenced Database Development and Regional Traffic Analysis | Dates: 01/11/2023 - 31/10/2024
Tools and Libraries: Python, NumPy, Scikit-learn, TensorFlow, Keras, Pandas, Matplotlib, SciPy

- Analysed the distribution of over 50,000 businesses across all counties in Ireland, covering the top 10 Nomenclature of Economic Activities (NACE) classes and key sectors such as Agriculture, Manufacturing, Construction, Wholesale, and Accommodation, identifying Agriculture, Forestry, and Fishing as the dominant category nationally.
- Revealed regional trends, including Wholesale and Retail Trade leading in Dublin, Accommodation and Food Services in Cork, and Agriculture, Forestry, and Fishing as a key sector in Galway and Limerick.
- Cork leads in agriculture businesses, with over 10,000 businesses, while Dublin dominates manufacturing and wholesale sectors with 1,000+ and 4,500+ businesses, respectively. Galway ranks high in accommodation services, with 3,000+ businesses, indicating its importance in tourism.
- Meath and Wexford show notable contributions to construction businesses, each exceeding 800 businesses, reflecting regional growth in infrastructure development.
- Limerick and Donegal emerge as significant hubs for diverse industries, including agriculture, wholesale, and accommodation, with multi-sectoral distributions exceeding 1,500 businesses in each category.

Project: A Neural Meta Model for Predicting Winter Wheat Crop Yield | Dates: 05/06/2023 - 31/07/2023
Tools and Libraries: Python, NumPy, Scikit-learn, TensorFlow, Keras, Pandas, Matplotlib, SciPy

- Designed an innovative multi-modal data solution integrating temporal and static datasets (weather and soil data), resulting in a 41% improvement in crop yield prediction accuracy compared to classical models.
- Developed a scalable architecture capable of processing 5 million+ records, maintaining near-linear memory and time efficiency.
- Published results in Machine Learning Journal.

Project: A Deep Learning Model for Heterogeneous Dataset Analysis | Dates: 30/03/2022 - 01/03/2023
Tools and Libraries: Python, NumPy, Scikit-learn, TensorFlow, Keras, Pandas, Matplotlib, SciPy

- Achieved an 11% improvement in prediction accuracy over classical models, validating on datasets with 1M+ records.

Project: Winter Wheat Crop Yield Prediction on Multiple Heterogeneous Datasets | Dates: 01/09/2018 - 16/12/2022
Tools and Libraries: Python, NumPy, Scikit-learn, Pandas, Matplotlib

- Increased reliability of predictions by integrating 2 diverse datasets, validated over 5 years of historical data.

Project: Sarcasm Detection in tweets | Dates: 01/09/2017 - 31/08/2018
Tools and Libraries: Python, Pandas, Scikit-learn, NumPy, TensorFlow/Keras, Matplotlib, NLTK

- Analysed over 479,000 tweets, performing comprehensive preprocessing to clean data for model training.
- Traditional ML models, such as Naive Bayes, achieved 79.05% accuracy, outperforming the provided SemEval benchmark system accuracy of 62.63% by over 16%.
- Deep learning models, including LSTM, achieved 77.29% accuracy with Glove embeddings, outperforming CNN (73.27%) and MLP (50%), demonstrating its suitability for sarcasm detection tasks