# Dr. Yogesh Bansal

Ireland

📞 +353-894780890 ✉ yogesh.bansal@ucdconnect.ie 🔗 https://www.linkedin.com/in/dr-yogeshbansal

⭘ https://dryogibansal.github.io/portfolio

## PROFILE

A results-oriented ML engineer and data analyst with extensive industrial experience working on real-world, comprehensive datasets. I have developed novel, cutting-edge machine learning models that are entirely unique and unparalleled in existing libraries, solving complex problems with innovative approaches. Skilled in building scalable, end-to-end data and ML solutions, I bring transferable expertise in data analysis, model development, and collaboration to drive impactful, data-informed strategies. Ready to contribute my unique skill set and proven ability to deliver novel ML solutions to Intercom.

## SKILLS

- **Data Science & Analytics:** Machine Learning, Data Analysis, Data Engineering, Cloud Data Engineering, TensorFlow, Keras, Numpy, Deep Learning, Data Mining, Statistics, Full Stack Development
- **Programming & Deployment:** Python, Flask APIs
- **Cloud & Big Data:** AWS, Hadoop, Spark, MapReduce, Amazon S3
- **Databases & Visualization:** MySQL, PostgreSQL, Pandas, Matplotlib, Seaborn, Tableau
- **Version Control:** Git
- **Project & Collaboration:** Production ML model development, Project management, Cross-functional collaboration

## PROFESSIONAL EXPERIENCE

### University College Dublin    Nov 2023 - Oct 2024
*Postdoctoral Fellow (Industry Collaboration with €100,000 Irish Government Dataset)*
- Performed extensive data preprocessing using ETL practices to ensure data consistency and establish reliable data analysis models, enhancing data-driven decision-making.
- Analyzed business trends and patterns, delivering actionable insights and personalized reports for stakeholders using tools like Tableau for visualization.
- Designed and implemented a georeferenced database with over 2 million records, integrating datasets for advanced regional analysis, utilizing SQL for data querying.
- Mentored junior researchers in data integration workflows, enhancing the team's proficiency in business-oriented data analysis.

### University College Dublin    Sep 2018 - Dec 2023
*Demonstrator and Teaching Assistant*
- Delivered training to over 200 students annually on machine learning, big data technologies, and enterprise business intelligence tools like Tableau, equipping them with strong analytical skills.
- Guided student teams in developing project proposals by translating ambiguous problems into structured data analysis scopes, fostering a practical understanding of data analysis.
- Linked academic programs with industry-leading technologies to prepare students effectively for data-driven roles with an emphasis on analytical skills.

### Init Call Technologies    Dec 2015 - Aug 2017
*Big Data Developer and Trainer*
- Optimized distributed data pipelines, reconfiguring MapReduce workflows to reduce query processing times from hours to under one hour for large-scale datasets.
- Conducted training sessions for teams on big data technologies, enabling them to efficiently process and analyze massive datasets.

### Irish Wheelchair Association    Jun 2018 - Aug 2018
*Information and Communication Technology Intern*
- Trained staff in modern ICT tools for data management, enhancing their understanding of data accuracy and consistency across operations.
- Implemented process improvements, reducing errors and increasing overall data reliability within the organization.

### Baddi University    Aug 2011 - Dec 2015
*Lecturer and Academic Mentor*
- Taught programming languages (C, C++, PHP) and data structures to over 200 undergraduate students, designing industry-aligned course modules.
- Supervised five postgraduate thesis projects, all successfully defended, focusing on industry-relevant technical challenges.

### Megh Softwares & DesignersX    Aug 2008 - Sep 2009
*Web Developer/PHP Programmer*
- Built and deployed dynamic websites for international clients, integrating content management systems like WordPress for enhanced usability.
- Enabled clients to independently manage website content, improving user engagement and operational efficiency.

## EDUCATIONAL ACHIEVEMENTS

**University College Dublin**                               **Sep 2018 - May 2024**

*PhD in Computer Science | Industry Collaboration with Origin Enterprises Using the Most Comprehensive Real-World Agricultural Dataset across UK & Ireland)*
- **GPA:** Passed
- **Achievements:** Conducted an industry-focused PhD in collaboration with Origin Enterprises, the largest agriculture tech company in the UK, innovating multi-modal data solutions by developing novel ML models.
- **Coursework:** Python, Applied Machine Learning, Cloud Computing, Data Mining, Database Development, SQL, PostgreSQL, Artificial Neural Networks, LSTM, Data Science, Data Analytics, Data-driven Decision Making, Visualization, Geospatial Data, Time Series Fore-casting, Classification, Statistics, Regression Models, Amazon Web Services (AWS), Analytical Skills, Academic Research, Optimization, Problem Solving.

**Dublin City University**                               **Sep 2017 - Aug 2018**

*Master of Science, Computing*
- **GPA:** 2nd Class Honours Grade 1
- **Achievements:** Developed deep learning models for sarcasm detection, yielding promising results.
- **Coursework:** Python, Applied Machine Learning, Cloud Computing, MySQL, Artificial Neural Networks, LSTM, Data Analytics, Visualization, Classification, Regression Models, Analytical Skills.

**Panjab University**                               **Oct 2009 - Jul 2011**

*Master of Engineering, IT*
- **GPA:** 1st Division
- **Achievements:** Designed a novel model for secure MANET communication, improving network stability and message transmission efficiency.
- **Coursework:** Mobile Adhoc Networks, Data Mining, Data warehousing, Algorithm Analysis and Design, Object oriented Analysis and Design

**Technological Institute of Textile and Sciences**                **Aug 2004 - Jul 2008**

*Bachelor of Technology, IT*
- **GPA:** 1st Division with Honours
- **Achievements:** Created software to digitize hospital operations, streamlining administrative workflows and enhancing efficiency.
- **Coursework:** C++, Java, Data Structures, Database Management, Operating System

## PROJECTS (MACHINE LEARNING)

*End-to-End Financial Risk Analysis Pipeline*                     **Nov 2024 - Present**
- Problem: Financial institutions face challenges in identifying high-risk profiles due to the complexity and scale of operational financial data, leading to suboptimal risk management strategies.
- Approach: Currently developing a pipeline leveraging ML models to analyse financial data and predict high-risk profiles. The pipeline employs advanced preprocessing techniques, feature engineering, and ML model development using PyTorch. The system is being built with a modular design for scalability, leveraging AWS Lambda for deployment and Docker for containerisation.

*Georeferenced Database Development and Regional Traffic Analysis*         **Nov 2023 - Oct 2024**
- Problem: Lack of comprehensive tools for understanding regional business trends and traffic patterns using large-scale georeferenced datasets and traffic information.
- Approach: Conducted extensive data preprocessing, including handling missing values using domain-specific interpolation approaches,
to ensure data consistency and reliability. Integrated datasets covering weather, traffic, and geodatabase information to develop a georeferenced database. Analysed the distribution of 50,000+ businesses across Ireland, covering key sectors such as Agriculture, Manufacturing, and Tourism.
- Results: Identified regional trends such as agriculture's dominance in Cork and tourism's significance in Galway. Findings supported regional planning and economic development initiatives.

*A Neural Mote Model for Predicting Winter Wheat Crop Yield*            **Jun 2020 - Aug 2023**
- Problem: Agricultural stakeholders lacked accurate ML models for predicting winter wheat crop yield due to the multi-modal nature of real-world datasets combining temporal weather data and static soil information.
- Approach: Conducted industrial-focused research in collaboration with Origin Enterprises, utilising real-world data from their extensive
customer base (thousands of clients). Developed the novel error-stabilised LSTM model (ESLSTM-MTF) for mixed temporal frequencies data to address these challenges. Performed scalability tests on datasets exceeding 5 million records, maintaining near-linear memory and time efficiency.
- Results: Achieved a 41% improvement in prediction accuracy compared to classical models. Published findings in the Machine Learning Journal, 2024, showcasing the practical applicability of the models for industry use.

*A Novel Deep Learning Model for Heterogeneous Dataset Analysis*          **Jun 2020 - Mar 2023**
- Problem: Traditional ML models struggled with the complexities of analyzing large, heterogeneous datasets used in agricultural forecasting.

- Approach: Proposed a variant of LSTM architecture to design and develop a novel deep learning model to handle multi-source datasets having differing temporal frequencies. Validated performance using over 1 million records from diverse input types, integrating weather and soil data for accurate crop yield prediction.
- Results: Improved prediction accuracy by 11% over classical models. Demonstrated scalability and robustness for industrial data analysis, contributing to data-driven agricultural strategies.

*Winter Wheat Crop Yield Prediction on Heterogeneous Datasets*                    **Sep 2018 - Dec 2022**
- Problem: Accurate crop yield prediction required effective integration of diverse datasets over extended timeframes.
- Approach: Utilised ensemble ML models to analyse five years of historical data, integrating weather and soil datasets to improve forecasting reliability.
- Results: Enhanced model performance and established the groundwork for the ESLSTM-MTF approach.

*Sarcasm Detection in tweets*                    **Sep 2017 - Aug 2018**
- Problem: Detecting sarcasm in textual data is challenging due to its reliance on contextual and subtle linguistic intricacies, which Traditional ML models often fail to capture effectively.
- Approach: Preprocessed over 479,000 tweets by cleaning data, handling imbalanced classes, and extracting features using natural language processing techniques. Implemented traditional ML models (Naive Bayes, SVM) alongside deep learning architectures (LSTM with Glove embeddings) to evaluate and compare performance, focusing on identifying sarcasm in tweets with improved accuracy.
- Results: Traditional ML models, such as Naive Bayes, achieved 79.05% accuracy and outperformed 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%).