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.

Education

University College Dublin, Ireland | Philosphy of Doctorate (Machine Learning) September 2018 – May 2024
Thesis Title**:** Enhancing Winter Wheat Crop Yield Predictions: A Data-Driven, Incremental, and Integrative Approach with ML.
Solved multi-modality data problem by designing and implementing innovative ML model architectures.
Designed and implemented novel ML models (**error-stabilised LSTM, weighted models**) for predictive analytics in agriculture.
Conducted scalability tests on large datasets, validating the models for real-world applications.
Key Achievements: Improved prediction accuracy by 41% using innovative architectures and scaled for large-scale deployment.

Dublin College University, Ireland | Masters in Data Analytics (Machine Learning) September 2017 – August 2018
Research Project Title: Deep learning-based sarcasm detection in tweets using NLP techniques.
Developed deep learning models for sarcasm detection in social media, achieving 70% accuracy.
Key Achievements: Published work in sentiment analysis and demonstrated expertise in TensorFlow, Python, and NLP techniques.
Relevant Coursework: Data Analytics, Visualisation, ML Algorithms, Deep Learning, Cloud Computing.

Panjab University, India | Masters of Engineering (Information Technology) October 2009 – July 2011
Maharshi Dayanand University, India | Bachelors (Information Technology) August 2004 – July 2008

Skills

Languages: Python, HTML/CSS, Javascript, Bash
Databases: MySQL, PostgreSQL, MongoDB
Libraries: NumPy, Pandas, Seaborn, Matplotlib
Frameworks: Flask, Keras, TensorFlow, PyTorch
Tools, Technologies & Cloud Platforms: Git, Docker, AWS Sagemaker, Airflow, CI/CD (Github Actions)

Experience

UCD, Ireland November 2024 – Present
ML Engineer (Freelance)
- Working on designing and implementing an end-to-end ML pipeline for financial risk analysis.
- This includes data gathering, preprocessing, feature engineering, model building, training, and deployment.

UCD, Ireland November 2023 – October 2024
Data Analyst and Project Lead (Postdoctoral Researcher)
- Led the development of a georeferenced database with 2M+ instances, integrating and analysing weather, geodirectory, and traffic datasets to derive actionable insights.
- Conducted advanced data analysis to identify regional business trends and traffic patterns, supporting strategic decision-making for regional development.
- Applied domain-specific interpolation techniques to manage missing data and delivered visualised findings.
- Mentored junior researchers in data integration methodologies, SQL, and Python-based analysis workflows.

UCD, Ireland September 2018 – December 2023
Technical Mentor and Project Facilitator
- Mentored and guided students on technical projects in Machine Learning, Deep Learning, Big Data, and Relational Databases.
- Designed and delivered practical lab sessions focused on implementing industry-relevant ML and Big Data solutions, including Hadoop, MapReduce, and Python-based pipelines.
- Supported the development of technical assignments and hands-on learning among diverse student groups.
- Acted as a project facilitator for multiple student teams for multiple projects.

Irish Wheelchair Association, Ireland June 2018 – August 2023
Information and Communication Technology Intern
- Conducted an evaluation of data systems and proposed strategies to improve data integrity.
- Recommended system improvements that enhanced data management practices.

Init Call Technologies, India December 2015 – August 2017
Big Data Developer and Trainer
- Optimised large-scale data pipelines using MapReduce and Hadoop.
- Deployed big data models on distributed systems and conducted team training on big data methodologies.

Baddi University, India August 2011 – December 2015

Technical Mentor (Assistant Professor)

- Led programming and data structure courses for undergraduate students.
- Supervised postgraduate students on technical thesis projects, specialising in mobile ad hoc network analysis.
- Collaborated with academic and technical teams to design and deliver curriculum aligned with industry in computer science.

Megh Software Pvt. Ltd. | DesignersX, India August 2008 – September 2009

Full-Stack Web Developer

- Designed and developed dynamic websites for overseas clients in Australia and America using PHP, JavaScript, AJAX, and CMS platforms like WordPress, delivered end-to-end full-stack web solutions.
- Implemented front-end interfaces, login scripts, data validations, and back-end functionalities.

Machine Learning Projects

End-to-End Financial Risk Analysis Pipeline November 2024 – Present

Tech Stack: Python, Pandas, Jenkins, Docker, AWS Lambda

- Currently working on designing and implementing an end-to-end ML pipeline for financial risk analysis including data gathering, preprocessing, feature engineering, model training, and deployment.

Georeferenced Database Development and Regional Traffic Analysis November 2023 – October 2024

Tech Stack: Python, Pandas, GIS, Matplotlib, SQL

- Designed and developed a georeferenced database with 2M+ instances, integrating weather, geodirectory, and traffic data to support data-driven decision-making.
- Dealt with missing values by applying domain specific interpolation approaches. Assessed the distributional count and share of businesses in each category for each county, identifying the top 10 business categories for regional development strategies.
- Mapped traffic patterns, focusing on volume, High Goods Vehicles, and Low Goods Vehicles concentrations.
- Utilised advanced data analysis and visualisation techniques to uncover regional business trends and transport dynamics.
- **Key Outcomes**: Delivered actionable insights for regional planning and transport infrastructure strategies.

Innovative ML Architectures for Multi-Modal Data: Crop Yield Prediction September 2020 – May 2024

Tech Stack: Python, Pandas, Matplotlib, Seaborn, TensorFlow, Scikit-learn, PostgreSQL

- Designed and implemented novel ML model architectures, which included an error-stabilised LSTM, weighted ML models and neural meta ML model, to address multi-modality data problems in agricultural winter wheat crop yield prediction.
- Conducted scalability tests to validate the models for real-world applications.
- **Key Outcomes**: Achieved a 41% improvement in prediction accuracy and demonstrated scalability for large-scale deployments.

Novel Contributions (Machine Learning)

A Neural Meta Model for Predicting Winter Wheat Crop Yield January 2024

- Designed a novel ML architecture, including an error-stabilized LSTM model to handle mixed temporal frequency data and a weighted ML model for extreme yield values.
- Combined these models into a neural meta model, leveraging their complementary strengths for enhanced predictions.
- Implemented, tested, and validated the models on real-world datasets, achieving a 41% improvement in accuracy over traditional approaches.
- Demonstrated scalability and efficiency with large datasets, enabling practical deployment for agricultural decision-making.

A Deep Learning Model for Heterogeneous Dataset Analysis May 2023

- Proposed and implemented an LSTM-based deep learning model for analysing heterogeneous datasets in crop yield prediction.
- Demonstrated a 11% improvement in accuracy over classical models, showcasing the benefits of leveraging time-series models in predictive analytics.

Winter Wheat Crop Yield Prediction Using Machine Learning December 2022

- Highlighted the benefits of integrating soil and weather datasets for crop yield prediction using machine learning models.
- Validated the effectiveness of combining static and temporal data, leading to improved prediction reliability and applicability in agricultural practices.

Deep learning-based sarcasm detection in tweets using NLP techniques August 2018

- Implemented both classical ML and deep learning approaches for sarcasm detection in tweets.
- Classical ML models initially outperformed deep learning models on a smaller dataset. However, with an increase in training data, the error gap in deep learning models reduced more than in classical ML models.