Contact

New Delhi
+353 830142910 (Home)
dhiraj.singh.xai@gmail.com

www.linkedin.com/in/dhiraj-singh-b6a8a51b7 (LinkedIn)

Top Skills

OpenVINO

Assistant Teaching

University Teaching

Publications

Forecasting cyber security threats landscape and associated technical trends in telehealth using Bidirectional Encoder Representations from Transformers (BERT)

Dhiraj Singh

PhD @ CRT in AI | Explainable AI | Senior Data Scientist at Edifecs | Ex-Hughes

Limerick, County Limerick, Ireland

Summary

A curious and passionate Data Scientist with the hunger to make an impact and who takes pride in building models that translate data points into business insights. Skilled in Exploratory Data Analysis and Visualization, Predictive Modeling, Machine Learning, Time Series Forecasting, A/B testing, Computer Vision, Natural Language Processing, Speech processing and SQL with a knack for narrating data-driven stories and actionable insights.

Furthermore, with a research background in multidisciplinary computational and physical fields of science, my research interests include open, emerging technologies such as artificial intelligence, blockchain, and augmented reality, as well as their management and technological forecasting. My primary research focus in AI domain includes Multimodal networks and Explainable AI.

Editorial Roles:

IEEE Computer Supported Cooperative Work in Design (CSCWD) 2022, Hangzhou, China
• Conference Member
• Organizing Chair Member for special session: Business Intelligence: The CSCW Perspectives

iCARTI 2021, Mauritius
• Committee Member, International Conference on Artificial Intelligence and its Applications

Research Reviewer:

• Multimedia Tools and Applications, Springer (2022)
• World Patent Information Journal, Elsevier (2021)
• Technological Forecasting & Social Change, Elsevier (2021)
• IEEE Transactions on Systems, Man, and Cybernetics

Experience

SFI Centre for Research Training in Artificial Intelligence

Doctoral Researcher

November 2023 - Present (10 months)

Limerick, County Limerick, Ireland

• Proficient in utilizing advanced AI-based methods to analyze and extract insights from complex multimodal data.
• Experienced in designing experiments and curating diverse datasets to establish robust research foundations, particularly in collaboration with Small and Medium Enterprises to address their evolving technological requirements.
• Actively involved in teaching assistant roles, providing guidance to peers, and offering support to professors.
• Integrates knowledge from diverse fields, promoting interdisciplinary approaches.
• Demonstrates creative problem solving skills by combining theory and practice to devise innovative solutions.
• Facilitates international collaborations and makes significant contributions to academic environments.
• Committed to shaping the future of Explainable AI, inspiring others through creative and simplified article writing in a Newsletter.
• Author of impactful papers, establishing a strong academic presence and contributing to research advancements.

Edifecs

2 years 6 months

Senior Data Scientist

July 2023 - October 2023 (4 months)

Chandigarh, India

• Data Science Infrastructure orchestration and management.
• Analyzing different sources of big data from the company for business value and process optimization.
• Researching and updating the team with the current State of The Art techniques for the given use case.
• Scaling training and inference pipeline design.
• ML model design to deployment inclusive of explainability (XAI).

Data Science Infrastructure Orchestration

• DataHub : Data Hub was used for data discovery, observability, and federated governance. This allowed Edifecs' Data Science team to easily examine the various data sources and schema available across various Edifecs products and teams.

• MLflow : MLflow is an open-source platform to manage the ML lifecycle, including experimentation, reproducibility, deployment, and a central model registry. Deployed an MLflow docker on EC2 to allow the data science team to track various experiments and Models.

• Automated Preliminary Exploratory Data Analysis : D-Tale, Autoviz, sweetviz, and Pandas profiling were used to automate the preliminary EDA. With the basic EDA ready in a few lines of code, data scientists were able to dive deeper.

• Explainable AI : Explain-ability is one of the main barriers AI is facing nowadays regarding its practical implementation specifically in the Healthcare domain hence we Designed AI explain-ability guidelines for different types of models. In addition, a collection of Explain-ability packages was created to assist data scientists in describing various sorts of ML models.

Data Scientist

May 2021 - July 2023 (2 years 3 months)

Projects:

• Risk Adjustments and Analytics - Knowledge Integrated Neural Network (KINN),
[AWS EMR, Kafka, Scala, Pyspark, Tensorflow Distributed, scikit-learn]

Developed a fully scalable and distributed pipeline for detecting missing medical conditions (ICD/HCC codes) from patients' medical history (claim data) (Suspect Engine). The architecture combines a Machine Learning method based on an Autoencoder - Recommendation system with an expert knowledge approach based on a Clinical Rule Engine.

• Talix Coding Insight Improvement and Talix Taxonomy Visualization,
[AWS EC2, Tensorflow, Pytorch, Detectron 2, HuggingFace Transformers, Microsoft UniLM, Doctr, Docker, FastAPI]

Talix coding insight helps clinical coders to find diagnostic codes in patient medical literature (charts). Improvements were made in the following directions.

• Reduced OCR costs by migrating from Abby Fine to Doctr based in-house OCR.
• Detection and segmentation of form type (Q&A versus OMR-based form) utilizing Mask R-CNN and layout parser.
• Q&A Form data extraction using LayoutLM V2.
• Deployed Dockerized models with FastAPI endpoints.
• Generated medical codes (ICD-10) with using Multimodal network with label attention mechanism.

The Talix Taxonomy maps out over 1,000,000 health concepts and over 2,000,000 relationships, making it the most comprehensive map of healthcare concepts in existence today. Designed and built a system that allows clinicians to visualise complex tabular taxonomy data in graph form.

• Population Payment Management – Episode of Care Risk Score,
[AWS EMR, Pyspark, MLlib]

Developed a system to determine Risk Coefficient associated with multiple types of episodes of care (Pregnancy, Cancer etc.) using Statistical Models. Risk Coefficient captures the demographic and clinical markers that affect the cost of care. Iterative GLM (Gaussian with Identity link) with Wald test-based feature drop was used. The P-value used was below 0.10.

University of Shanghai for Science and Technology

2 years 9 months

Research Assistant

February 2021 - October 2023 (2 years 9 months)

Education in Web3: Understanding the Ed-Surge Current Status and Future Directions.
Prof. Usharani Hareesh Govindarajan
Dr Hong-Zheng Sun-Lin

Web3 in education has the potential to significantly enhance the accessibility, security, and authenticity of educational resources and credentials. This research presents the current education status in Web3, the driving factors, and the future directions and applications of Edu-verse in Web3.

Forecasting cyber security threats landscape and associated technical trends in tele-health using Bidirectional Encoder Representations from Transformers (BERT)
Prof. Usharani Hareesh Govindarajan
Prof. Hardik Gohel

The research presents a systematic visualisation across a large corpus of global patent grants from over 40 key patent offices for a better understanding of the cybersecurity technology landscape and emerging trends for Telehealth.

Blockchain Technologies Enabling Business & Enterprise Management: Key Technology Cluster And Ownership Trends,
Prof. Usharani Hareesh Govindarajan
Prof. Dali Zhang

This research uses a collaborative human and machine intelligence approach to train on a global patent grants dataset containing 37,884 patents towards the identification of key technology clusters, and their ownership trends for blockchain technologies applied in the business and enterprise management related area.

Team: Xiaojun Lu, Dali Zhang, Dhiraj Kumar Singh, Usharani Hareesh Govindarajan, ChongJun Fan.

Emerging Design Informatics Directions Aiding Citizen Development,
Prof. Usharani Hareesh Govindarajan

The research suggests a plausible solution that proposes and enables business managers to prototype an automated analytics pipeline through the proposed middle-out design approach using quick, readily available low-code tools and foundational concepts to acquire a working vocabulary for communicating with data science teams.

Teaching Assistant
January 2023 - April 2023 (4 months)
Shanghai, China

The content covered in the "Data Mining" course encompasses various phases of the data mining process, including problem definition, analytics design, data exploration, data preparation, modeling using open-source AI platforms and Python programming, and visualization. The lectures aim to establish a practical groundwork for real-world applications, incorporating case study discussions within the classroom.

Teaching Assistant
September 2022 - December 2022 (4 months)
Shanghai, China

This course titled "Python programming" is intended to lay the groundwork for real-world applications, which will be followed by an advanced Python programming course titled "Python Applications" in the final semester of the undergraduate program.

SKEMA Business School
Teaching Assistant
September 2021 - December 2021 (4 months)
Suzhou, Jiangsu, China

The Master's in Science (International Business) program, titled "Artificial Intelligence for International Business," introduces the transformative aspects of Artificial Intelligence (AI) technologies in the context of business. This course adopts a data-driven approach, underscoring the pivotal role of data as the foundation for industry projects. Despite the potent possibilities AI offers for business transformation, its intricate complexity often makes it challenging for individuals with diverse subject matter expertise to fully grasp its potential.

In response to this gap, the course delineates the comprehension, design, and utilization of business analytics using low-code tools like the Microsoft Power Platform. Its primary aim is to equip future International Business professionals with AI terminology, positioning them as citizen developers capable of enhancing their teams' and organizations' performance. Additionally, the course serves as a practical reference guide for the Exam PL-900: Microsoft Power Platform Fundamentals professional certification exams.

Hughes Systique Corporation (HSC)
Data Scientist
July 2019 - May 2021 (1 year 11 months)

Business problem understanding and optimization
Dataset exploration, visualization, Statistical Analysis, and Feature Engineering.
Model training, evaluation, and testing.
Improving predictions with hyper-parameter optimization.

Projects:

Inventory Management System for Commercial Refrigerators
• Leveraged Monocular Depth Estimation and Object Detection for Coca-Cola India.
• Achieved 98% accuracy in determining bottle counts for efficient inventory management.
• Applied Data Analysis and Time Series Forecasting to predict usage trends and seasonality.

Measuring Retail Store Traffic System
• Implemented Person Re-Identification Algorithm, achieving a 71% Multiple Object Tracking Accuracy.
• Tracked store visitors in real-time, optimizing store space management.
• Developed a scalable architecture based on distributed computation for efficiency.

Social Distancing and COVID-19 Management
• Engineered a system using projective transform and camera calibration to enforce social distancing.
• Created an analytic dashboard for administrators to identify areas and times prone to violations.

Augmented Reality-Based Installer Bot
• Designed an augmented reality system guiding users through network device setup.
• Used object detection, OpenGL, and OpenCV for accurate device orientation determination.

Image Installation Auditor Tool
• Automated auditing processes, achieving a 95% efficiency boost.
• Implemented image classification and segmentation for accurate equipment classification.

Call Centre Transcripts Generation

• Utilized speech signal processing and speaker diarization for a 25% Word Error Rate on real-world audio.

• Outperformed Google's Speech API accuracy by 5% through precise speech-to-text conversion.

Natural Language Understanding and Time Series Prediction

• Developed a model with a 92% accuracy rate in predicting customer chat behaviour.

• Enabled proactive measures to retain customers at risk of opting out of services.

IIC research lab

Data Science Intern

June 2018 - November 2018 (6 months)

Delhi

worked on Arrhythmia Detection and Management using IOT and Deep learning

---

Education

University of Limerick

Doctor of Philosophy - PhD, Artificial Intelligence · (November 2023 - October 2027)

University Grants Commission of India

National Eligibility Test (UGC-NET), Computer Science · (October 2019 - October 2019)

IIC

Master's degree, Informatics and communications · (2017 - 2019)

Christ University, Bangalore

Bachelor's degree, Triple honour - Physics mathematics Electronics · (2013 - 2017)