HARSH DHINGRA

dhingrah02@gmail.com |+918448779805 | https://www.linkedin.com/in/harsh-dhingra/

EDUCATION

Bachelors in Engineering (Computer Engineering) ***2017 - 2021***
Thapar Institute of Engineering and Technology & Trinity College Dublin *GPA|(3.54/4)*
Coursework:
Deep Learning, Internet Applications, Security and Privacy, Digital Image Processing, Microprocessors, Software Design Analysis, Information Management, Digital System Design, Management for Engineers, Data Structures, Optimization Techniques, Engineering Design, and Operating Systems.

WORK EXPERIENCE

DATA SCIENTIST | ***Lighthouse Global, Bangalore*** ***Jul 2022-Present***

- Architected advanced ML solutions for eDiscovery products using Python frameworks, optimizing data processing and analysis.
- Improved ML model accuracy by researching and analyzing algorithms (BIRCH, Neural Network, Vector Space Model, Clustering, SVM Classification) transitioning to the transformative BERT model, further elevating the performance and capabilities of the system
- Designed a scalable ML Pipeline on Azure Machine Learning Studio, classifying over a million documents efficiently.
- Successfully managed project flow and provided crucial support for cloud pipelines, guaranteeing seamless operation of various products.
- Transformed codebase, reducing HSQL reliance and enhancing data storage with Delta Lake.
- CAL eDiscovery Project: Spearheaded the Continuous Active Learning (CAL) eDiscovery project, incorporating advanced techniques and leveraging DistilBERT Transformers to elevate the document processing and categorization process.
- Mentored and supervised two interns, enhancing their professional growth. Implemented pytest testing, raising test coverage from 20% to 64%. Collaborated on an Azure Data Factory pipeline for comprehensive product unification, covering machine learning pipelines, preprocessing, model preparation, and scoring.

Project: Azure Kubernetes Spark Integration (AKS on Spark)

- Implemented AKS on Spark to run ML models, reducing dependency on Azure Databricks (ADB).
- Achieved enhanced flexibility and control over Spark jobs.
- Significantly reduced operational costs.
- Utilized tech stack: Spark, Python, Azure Machine Learning.

Optimising Databases and Migrating to Delta Lake

- Conducted a Proof of Concept (POC) to evaluate alternatives to the slow and expensive HSQL database. Tested various file types on a 10k-document dataset to assess read-write performance.
- Demonstrated that Delta Lake outperformed other options and presented findings to senior management.
- Transitioned to production by leading the migration of existing databases and workflows from HSQL to Delta Lake.
- Managed the migration using Python, PySpark, Spark, Azure Data Factory, Azure Databricks, and Azure Blob Storage.
- Implemented cost optimization strategies in storage during the migration process.

ACHIEVEMENTS AND GROWTH

Topic Modelling in Lighthouse Hackathon:

- Developed a Topic Modeling solution as part of the Lighthouse Hackathon and successfully presented the solution to stakeholders
- Conducted a proof of concept (POC) using a sample dataset containing one million data points.
- Implemented Latent Dirichlet Allocation (LDA) on Azure Databricks.
- Utilised algorithms such as LDA, Python, and Spark for this project.

Generative AI for Key Document Identification and Text Generation

- Attained a 95% accuracy rate in automated document identification and summarisation utilizing GPT-3.
- Designed and implemented bespoke document identification solutions for diverse clients.
- Successfully certified in Generative AI Fundamentals by Databricks. *(here)*

- Implemented advanced machine learning models using Python, TensorFlow, and PyTorch, alongside GPT-3 for enhanced natural language processing. Integrated Azure Open AI service for additional functionality.

ENGINEER-1 | *Samsung SDS* **Jul 2021 - July 2022**

Samsung Promotool

- Created training and testing datasets by extracting customer web session data, facilitating feature selection for Machine Learning models to predict customer behaviour (abort/continue) in product promotion campaigns.
- Implemented data collection modules within the internally developed Campaign Review System, enabling efficient data acquisition from web application clickstream and purchase data sources

Samsung Knox

- Enhanced user-facing functionality across all three tiers of the Samsung Knox application, contributing to its overall usability and effectiveness.
- Added comprehensive tests and diligently addressed bugs, ensuring the stability and reliability of the Samsung Knox application.

ACADEMIC PROJECTS

Assessing skill from videos - Next step towards automatic skill evaluation (Research Thesis)

- Created an efficient deep-ranking neural network model, employing TSN and 2SCNN to analyse the interplay of temporal and spatial dimensions. Implemented a beta-driven single loss function, achieving an 85% reduction in computational overhead compared to the reference paper.
- Leveraged expertise in Convolutional Neural Networks (CNNs), Temporal Segment Networks (TSN), and Two-Stream Convolutional Neural Networks. Employed a Siamese architecture with ALEXNET to design a dual-stream model—one stream focusing on spatial information for image training and the other stream handling temporal aspects.
- Utilised K-Fold Cross Validation for robust model evaluation. Proficiently worked with libraries including NUMPY, PANDAS, SCIKIT-LEARN, PYDOT, TENSORFLOW-GPU, and CONDA.

Prognosticate Sentiment Emotions from Social Media Data: Illuminating the Societal Pulse

- Proficient in advanced NLP processing techniques for emotion categorization, achieving a 94% accuracy with LSTM Neural Networks.
- Developed a robust analytical mindset, evaluating models like Decision Tree and Random Forest. Applied meticulous data pe-processing for accurate feature extraction.
- Managed extensive datasets using techniques such as data cleaning, tokenization, and word embeddings, showcasing effective pre-processing of ML
- Introduced a framework prioritising contextual awareness, paving the way for future sentient analysis advancements
- Successfully applied the model to predict emotions in tweets, serving as a valuable tool for early depression detection – a substantial contribution to mental health, showcasing practical significance and societal impact.

Spam Email Detector

- Designed a robust machine learning solution employing NLP techniques, Naïve Bayes, and SVM algorithms to categorise emails as "spam" or "ham efficiently."
- Conducted meticulous pre-processing using the NLP toolkit on a dataset encompassing 10,000 emails.
- Achieved impressive classification accuracy rates of 95% for Naïve Bayes and 93% for SVM, corroborated by a comprehensive confusion matrix analysis.

Machine Learning in Trading: A Comprehensive Approach

- Developed and trained diverse machine learning models, including support vector machines and deep learning architectures, to identify profitable pair trading opportunities.
- Utilized historical data to instruct models in understanding correlations and patterns, leading to accurate trading signals for identified pair trading opportunities.
- Predicted Currency Trends: Employed supervised machine learning, particularly Support Vector Machine, to forecast currency rate trends.
- Applied various classification and regression algorithms to predict trends accurately. Employed regression techniques to compute precise daily, weekly, and monthly trading ranges.
- Orchestrated components for effective trading decision-making.

COURSES

- Computer Vision Fundamentals with Watson and OpenCV, IBM
- Traffic Sign Classification using Deep Learning in Python/Keras, Course-era guided project.
- Machine Learning with Python, IBM
- Neural Networks and Deep Learning, deeplearning.ai
- Databricks Certification *(Scalable Machine Learning with Apache Spark , generative AI fundamentals)*
- Introduction to Data Science in Python, University of Michigan
- Support Vector Machines with sci-kit-learn. Coursera Project Network
- Software Processes and Agile Practices
- Python Data Structures

SKILLS

- **Programming:** Python, Java, SQL, HTML/CSS/JavaScript
- **Frameworks and Tools:** Keras, TensorFlow, Git, Jupyter, Reactjs, PyCharm, VS Code
- **Software Engineering:** Agile, Team Development
- **Cloud Tools:** Azure Machine Learning Studio, Databricks, Azure Kubernetes Service, Azure Data Factory, Azure Data Studio, Azure Open AI