Winnie Winnie

linkedin.com/in/winniewi

Data Scientist | ML Engineer | AI Innovator | Full Stack

SKILLS / TOOLS

Core

Python | Java | SQL

Data Science / ML

Pandas | Sklearn | Numpy | Jax
Tensorflow | Keras | PyTorch | PyG

LLM

OpenAI | Langchain | HuggingFace
VectorDB | RAG | Gemini | LLaMA

Data Processing

ApacheBeam | PySpark | Ray | Dask

Data Visualisation

Matplotlib | Seaborn | Plotly | Looker

Other Tools

Airflow | Conda | Docker | Git

API / Web

Flask | Fast | JS/TS | React | Angular

Cloud Services

GCP: BigQuery | Colab | VertexAI
Cloud Storage | Dataproc | Dataflow
AWS: Redshift | Beanstalk | Snowflake
ElasticSearch | Dynamo | S3 | Lambda

EDUCATION

Sheridan Institute of Technology
Computer Programmer (2yr)
Silver Medalist (Top of Class)

York University
BSc Physics & Astronomy
First Class with Distinction

EXPERIENCE

Pelmorex Weather Solutions (The Weather network) Oakville, ON
DATA SCIENTIST Mar 2020 – Current

Machine Learning

- Developed **time series models** to forecast user behavior based on weather & seasonal patterns, improving demand prediction for inventory management
- Optimized **distributed model training** for 100k+ (upto 1M) models using Ray & Dataflow, reducing cloud cost estimates from ~$10k to ~$1.5k
- Organized an evaluation study for global vs localized forecasting models across geographic regions
- Worked with meteorologists to develop an **ensemble model** of multiple Numerical Weather Predictions, improving forecast accuracy and contributing to ForecastWatch leadership
- Collaborated on a **GNN encoder-decoder** model (Graphcast) for medium range weather forecasting, enhancing in house production forecasting models
- Built a **CNN-based** solution for extracting building geometries & classifying Point of Interest from geospatial imagery, reducing GIS workload & speeding client delivery
- Developed user **segmentation and clustering** models leveraging web browsing patterns, visit frequencies, durations, and physical location data to build detailed behavioral profiles and uncover actionable market trends
- Designed and deployed an **end-to-end ML pipeline** to predict user age and gender from online behavior data, increasing demographic coverage from 6% (ground truth) to over 80% of users
- Improved demographic prediction accuracy to 70% by implementing an ensemble with **majority voting** and weighted techniques
- Identified system and selection bias skewing the data & used SMOTE, batch sampling strategies, weighted models & weighted loss functions for **imbalanced data handling**
- **Applied ML** models like Random Forest, XGBoost, Logistic and Linear Regression, ANNs in production pipelines
- Employed strategies including PCA, correlation strategies like Pearson, chi2, mutual information, KS tests, confusion matrix, ROC curve & SHAP values to analyze the model & the data
- Developed **prediction intelligence** pipelines to monitor model performance, detect **data drift** ensuring continuous model optimization

Statistical Modeling & Analytics

- Correlated client data with weather patterns identifying **seasonality impacts** & **behaviour trends** to power a visual dashboard turning it into a revenue generating product
- Worked on **similiarity scores** based on user behaviour enhancing the statistical significance of insights for niche audiences, even with limited data
- Pioneered a **demographic insights** solution for targeted billboard advertising using **location and behavioral data** , bringing in 2 major billboard clients
- Engineered **dynamic population-scaling** multiplier to address device reach and data inconsistencies - giving 10% margin with quantifiable real-world data
- Developed performance metrics like **incremental visits & lift** resulting in a

RELEVANT COURSES

Object Oriented Programming
Database Design (SQL, NoSQL)
Systems Analysis & Design
Java Enterprise & C# .NET core
User Interface & UX design
Linux System Admin
Data Analysis
Multivariate Calculus
Linear Algebra
Differential Equations

Online

ML (Andrew Ng) - Coursera
Generative AI - Google
ML Engineer - Google
Graph NN CS224W - Stanford
Udemy:
Data Science A-Z
Deep Learning A-Z
LLM Engg. RAG & LoRA
AI A-Z (Deep Q-learning)

github.com/w-winnie
https://livnlearns.com/

data-driven campaign optimization strategy maximizing ROI

- Designed adaptive strategies for selecting **control groups** for **behaviour normalization** to account for unquantifiable factors biasing insights

LLM

- Leveraged **OpenAI GPT models** for summarization workflows in productionized client facing dashboards
- Developed a custom reporting tool using **few shots prompting, functional tool calling** and **agentic approach** - covering **text to sql** and visualization
- Incorporated **entity recognition** using open source models & transformers from Hugging Face, enhancing the accuracy of model response
- Built **RAG pipelines** integrating metadata and web content for contextual targeting based on online behaviour
- Familiar with techniques like TF-IDF, entity density scoring & syntactic parsing
- **Evaluated LLM pipelines** with metrics like BLEU, ROUGE, perplexity, cosine similarity, AI as a judge etc

FULL STACK DEVELOPER - DATA — Sep 2019 – Feb 2020

- Developed **RESTful APIs** & **ETL pipelines** for **real-time** metrics of geospatial data products & client dashboards
- **Audited** third party geo data providers **reducing** data acquisition **costs by 30%**
- Utilized Big Data tech for automated processing & **visualization** of **GBs** of data
- **End-to-end** metrics - from conceptualizing logic, updating data model, loaders, setting up ETL tasks, API endpoints to creating frontend visualizations

DATA INTERN — 4mo.

- Accomodated new data model requirements in workflows
- Helped automate ad campaign workflows with **metadata driven orchestration**
- Migrated monolithic APIs into **microservices** for **scalable geospatial** workflows
- Secured APIs against **SQL injection & data leaks** via best practices

Canadian Hydrogen Intensity Mapping Experiment (CHIME)

INTERN — 8mo.

- Analyzed Fast Radio Burst (FRB) catalogs recorded by CHIME, developing criteria to distinguish sidelobe vs mainlobe events in **radio telescope data**
- Created a **CNN classifier** for sidelobe event detection in **real-time (milliseconds)** , enabling outriggers to capture additional data in time
- Investigated Radio Frequency Interference (RFI) patterns & optimized models for **precision-recall** , providing mitigation strategies & false-positive estimates