GAGANDEEP SINGH

Brampton, ON L6Y 1N9
(437) 833-7766 | singh.gagandeep1103@gmail.com

DATA SCIENTIST | MACHINE LEARNING ENGINEER

Detail-oriented and analytical professional with a passion for harnessing the power of data to drive insightful business decisions. Adept at applying statistical techniques, data mining, and machine learning algorithms to extract meaningful insights and solve complex problems. Excellent problem-solving abilities combined with effective communication skills, enabling the translation of technical findings into actionable recommendations. Continuously seeking to expand knowledge and skills in the rapidly evolving field of data analytics.

CORE SKILLS & COMPETENCIES

- ML Algorithms
- AWS SageMaker
- Artifact Tracking
- A/B Testing
- Python Programming
- Data Visualization
- Hydra Configuration
- Database Management
- Experimental Design
- Generative AI
- NLP
- AWS Lambda

ACADEMIC ACCOMPLISHMENTS

Bachelor of Science in Data Analytic | Western Governor University | Salt Lake City, UT **March 2024**

Relevant Coursework: ML Algorithms | MLOps | Statistical Testing | Exploratory Data Analysis | Data Management – Foundations | Programming in Python | Scripting & Programming – Foundations | Data Structures and Algorithms I | Web Development Foundations | Artifact Tracking | Version Control Tools

Key Skills Acquired:

- ***Data Analysis:*** Acquired comprehensive knowledge and skills in data analysis techniques, including data cleaning, visualization, statistical analysis, and predictive modelling. Proficient in using Python, and data management tools to extract, transform, and analyze data.
- ***Statistical Modeling:*** Developed expertise in statistical modelling techniques, including regression analysis, hypothesis testing, and time series analysis. Skilled in applying statistical methods to identify patterns, trends, and relationships in data, enabling informed decision-making and actionable insights.
- ***Machine Learning:*** Developed a strong foundation in machine learning algorithms and techniques, including supervised and unsupervised learning, classification, clustering, and recommendation systems. Proficient in using libraries and to build and deploy machine learning models.
- ***ML Pipeline AND Artifact Tracking:*** Acquired foundational skills in setting up machine learning (ML) pipeline components by utilizing artifact tracking through tools such as Weights & Biases (W&B), MLflow, and Conda environments.

Key Projects:

ML Pipeline for Short-term Rental Prices in NYC

Developed an ML pipeline for estimating housing prices in New York City, incorporating various data components and utilizing Conda environments and ML tools for data extraction, exploratory analysis, data cleaning, and model training.

- Analyzed housing data using Python in Jupyter Notebook, conducting exploratory data analysis, preprocessing, and artifact tracking. Segmented the pipeline into distinct components for data analysis, cleaning, testing, splitting, model training, hyperparameter optimization, and visualization. Integrated weights and biases for performance monitoring.
- Successfully released the finalized pipeline on GitHub, and gained approval from the instructor. Achieved enhanced accuracy scores and created effective ML components using ML tools such as MLflow, contributing to accurate housing price predictions in New York City.

GAGANDEEP SINGH (437) 833-7766 | singh.gagandeep1103@gmail.com

ACADEMIC ACCOMPLISHMENTS (CONTINUED)

Identify Customer Segments Using Unsupervised Machine Learning

Utilized Python ML Libraries within a Conda environment and Jupyter Lab to identify customer segments for targeted marketing initiatives.

- Transformed categorical and mixed-type features through feature engineering, implemented feature scaling, and applied dimensionality reduction techniques like Principal Component Analysis (PCA). Utilized K-means clustering to identify and analyze customer segments, providing actionable insights for informed business decision-making.
- Identified shared households as a more likely customer group to purchase the company's products compared to pre-family couples and individuals. Aligned marketing initiatives to capitalize on the target market of shared households, improving marketing effectiveness and potential sales.

Conducted A/B Testing to Analyze the Implementation of a New Web Page

Implemented webpage analytics to track user engagement and behaviour on the digital platform. Utilized A/B testing methodology to compare performance metrics such as click rates and impressions between the existing webpage and the newer version.

- Led the execution of an A/B test, comparing the performance of two webpage variations. Employed a two-sample Z-test and carefully examined p-values to conduct hypothesis testing, revealing that the conversion rate of the newer page was significantly worse than that of the previous page.
- Analyzed the data and identified that the US and UK regions had a lower conversion rate. Recommended optimizing the existing webpage instead of creating a new one, as no benefits were found in introducing a new design.

PROFESSIONAL EXPERIENCE

ML Engineer | Vosyn AI | Contract — **Feb. 2024 - present**

Data Engineer and Web Developer | Leuto | Freelance — **Aug. 2019 – Dec. 2023**

Quality Assurance | Kingspan Insulated Panels | Caledon, ON — **Jan. 2019 – June 2021**

Team Lead, Machine Operator | Kingspan Insulated Panels | Caledon, ON — **Jan. 2018 – Jan. 2019**

CERTIFICATIONS

Advanced NLP and GEN AI | Udemy — **In Progress**

AWS Certified Cloud Practitioner | Amazon — **Dec. 2022**

CompTIA Data+ Certification | CompTIA — **Nov. 2023**

Udacity ML Ops Certification | Udacity — **Feb. 2024**

Udacity Data Analyst Nano Degree | Udacity — **Jan. 2024**

TECHNICAL ACUMEN

Software Skills	Python | C++ | Oracle | MongoDB | JavaScript | SQL | Tableau | Power BI | Excel | Jupyter Suite | Git
Libraries/Frameworks	NumPy | Pandas | Scikit-Learn | TensorFlow | Keras | PyTorch | Matplotlib | Flask | Django | Spark | NodeJs | ExpressJs
Data Analytics	Machine Learning | Deep Learning | Natural Language Processing (NLP) | Cloud Computing | Statistical Analysis | Generative AI