Sammy Elhindi, Computer Science/ Data Science

Dallas, TX; (214) 554 – 2370; selhindi10@gmail.com**; linkedin.com/in/sammyelhindi;** sammyelhindi.net

---

Professional Summary:

- **Seasoned Data Scientist with 2+ Years of Experience:** Certified by IBM in Data Science and Machine Learning. Expert in statistical methodologies, machine learning, cloud computing, and the forefront of Generative AI technologies, including Large Language Models

- **Expert in Data Analytics and Machine Learning:** Proficient in developing and deploying ML models in production environments. Expertise in NLP (Natural Language Processing) for enhanced data analysis and decision-making.

- **Advanced Skills in AI and NLP:** Strong foundation in mathematics and statistics, with hands-on experience in Artificial Intelligence, Machine Learning, NLP, and Predictive Modeling. Recently expanded expertise to include developing and optimizing Large Language Models (LLMs) such as GPT-3, GPT-3.5, GPT-4, and LLAMA (1-3) as well as tools like CoPilot for sophisticated NLP and AI applications.

- **Data Mining and Feature Engineering Specialist**: Extensive experience in data mining, data preparation, exploratory analysis, and feature engineering using supervised and unsupervised modeling techniques. Integrated Large Language Models (LLMs) to achieve more nuanced and precise analyses.

- **Comprehensive Data Handling and Visualization**: Skilled in creating impactful data visualizations and dashboards using Power BI and Tableau, incorporating LLM outputs to enhance reporting insights.

- **Strong Team Builder and Effective Communicator:** Known for excellent communication, time management, resource management, and client relationship development skills. Demonstrated ability to lead teams toward integrating cutting-edge technologies like LLMs into business processes.

Technical Expertise

- **Programming Languages**: Python (Pandas, Numpy, Sklearn, PyCaret), R (R-Studio)
- **Machine Learning**: Naïve Bayes, K-NN, SVM, Decision Tree, ExtraTrees, XGBoost, Random Forest, K-Means, AdaBoost, Linear/Multiple Linear Regression, SVR, Prophet
- **Natural Language Processing & Large Language Models**: NLTK, GPT-3, GPT-3.5, GPT-4, LLAMA (1-3), CoPilot
- **Data Visualization**: Power BI, Tableau
- **Databases**: SQL, MySQL, NoSQL
- **Tools**: Git, GitHub, Visual Studio Code
- Statistical Analysis & Mathematics

Professional Experiences

Junior Data Scientist   August 2023 – Present

Clients: Amazon; Zillow; Texas Independent School Districts; Healthcare Provider, TX;

Projects: *Sentiment analysis, Anomaly Detection, Price Forecasting, Patient Discharge Prediction, Heart Attack Prediction*

- Developed customized **Machine Learning** models tailored to Zillow, and Healthcare Provider clients, incorporating cutting-edge techniques such as **Neural Networks** and **ExtraTrees Algorithms,** optimizing prediction accuracy, and aligning models with clients' specific business goals.
- Utilized advanced **NLP** techniques for sentiment analysis and text mining, generating actionable insights that improved customer experience for financial services clients and supported data-driven decision-making.
- Implemented explainable **AI methods** to ensure transparency and interpretability of complex models, fostering trust with clients and enhancing model accountability.
- Conducted **A/B testing** and designed experimentation frameworks that enabled data-driven product optimization strategies, ensuring seamless user experiences and effective deployment of machine learning models.
- Collaborated with cross-functional teams to develop **AI-powered** chatbots and virtual assistants for healthcare clients, enhancing customer engagement and streamlining support processes.
- Designed and executed Supervised and Unsupervised Machine Learning experiments using Python, delivering detailed predictive analytics and developing models for continuous, nominal, and ordinal data to support diverse business needs.
- **Evaluated classification algorithms** using metrics such as **precision**, **recall**, **F1-score**, **confusion matrix**, **AUC** for precision-recall, and **ROC curves**, ensuring model robustness and improved predictive performance.
- Developed **anomaly detection** and **clustering techniques**, enabling precise outlier identification and enhancing data analysis for more accurate business insights and better decision-making.
- Led projects involving **Data Cleaning, Feature Engineering, and Machine Learning**, consistently delivering results on time, within budget, and aligned with client expectations.
- Assessed model fairness and bias using specialized metrics, ensuring ethical AI practices and mitigating biases in predictions, crucial for sensitive data projects in insurance and healthcare.
- **Utilized statistical measures** such as **R-squared**, **MAE**, **MAPE**, and p-value to evaluate models, extract key findings, and drive actionable insights for data-driven improvements.
- Created intuitive data visualizations using **Power BI** and **Tableau**, publishing interactive dashboards that streamlined reporting and enhanced decision-making for stakeholders.

Software Engineer   August 2022 – Present

Projects: *Evolutionary Algorithm to Develop Neural Networks, Computer Vision*

- Collaborated with a cross-functional team to design and implement an advanced growth-based evolutionary algorithm extending NEAT (NeuroEvolution of Augmenting Topologies), enabling scalable neural network architecture inspired by biological neuron development.
- Utilized Python to extract visual features, detect corners, and classify artwork from various images.

Education
- **BS in Computer Science -** The University of Texas at Dallas, December 2025
- Ranked 1st in a graduating class of 720, National Merit Commended Scholar, Top 1% SAT Score, AP Scholar with Distinction, First Place – TSA Regional & State Competitions in Debating Tech Issues.