Baofang (Luna) Zhang, PhD

Bear, DE 330.319.3863
www.linkedin.com/in/baofang-luna-zhang
https://github.com/zhangbf2020/project zhangbf159@gmail.com

Solve Complex Business Challenges by Gathering Key Insights Using Data Modeling, Data Visualization, and Data Science Methodologies

- Analytical and innovative, collaborating well across teams to deliver projects on time and within budget.
- Extract insights from data through data mining, data visualization, and machine learning (ML) to help business leaders make informed decisions. Bring new technology Generative AI, LLM to business, developing products for customers based on business requirements. Experience spans supply chain, retails, marketing, and healthcare.
- Proven expertise in managing processes and business methodologies and mentoring and leading teams to achieve goals.
- Customer-focused and adaptable to a variety of work environments, with high tolerance for risk and stress associated with strict deadlines.

Skills

- Python (NumPy, Pandas, Matplotlib, Seaborn, Sklearn, Scipy, Tensorflow, Pytorch, Keras, Requests, Beautiful soup), RNNs, NLP (NLTK), SpaCy, PySpark, Prompt Engineering, SQL (MySQL, PostgreSQL), R, Supply Chain, Business Analytics, Marketing.
- Regression, Classification, Clustering, Experiment Design and Analysis, Feature Engineering, Time Series, Statistical / Predictive Modeling, Segmentation, A/B testing, Bayesian inference, Linear Regression, Logistic Regression, Decision Tree, Random Forest, GBoost, XGBoost, SVM, Generative AI, LLM.
- Jupyter Notebook, Linux, GitHub, Jira, AWS, Flask, Google Cloud Platform(GCP), Slack, SSDM, Visual Studio.

Experience

DUPONT, Wilmington, DE 02/2021 – 01/2024

Data Scientist Lead / ML/AI Subject Matter Expert (SME)

Project Title: Advancing Chemical Research through NLP and LLM Integration

- Developed a comprehensive support system, which integrated Large Language Models (LLM) and Retrieval-Augmented Generation (RAG) to understand complex chemical queries, retrieve relevant scientific literature, and generate concise, insightful summaries and answers to facilitate research and development.
- Implemented Named Entity Recognition (NER) and Relationship Extraction (RE) models to identify chemical entities and their interactions accurately.
- Significantly reduced the time required for literature review and hypothesis generation in chemical research.

Project Title: Leveraging Machine Learning to Uncover Business Insights from Sales and Marketing Data

- Developed machine learning models to predict sales trends, and analyzed marketing indicators, providing insights that inform marketing strategies and business decisions; conducted **A/B testing** on selected marketing strategies to validate the models' recommendations; achieved a 10% increase in sales revenue attributed to optimized marketing strategies and a reduction in marketing costs by 15% without compromising campaign effectiveness.
- Led 5 demand analysts in forecasting monthly demands, using timeseries models for supply chain and financial groups, allowing ability to make production plans and optimize inventory.

TECHLENT, INC, San Mateo, CA 01/2020 – 01/2021

Data Scientist

Project Title: Intelligent Pricing Strategy for E-Commerce Enhancement

- Developed data-driven pricing strategies for various product categories within the e-commerce platform leveraging advanced analytics and machine learning algorithms.
- Utilized regression analysis to understand the impact of various factors on product pricing and sales performance.
- Achieved a significant increase in sales volume and profit margin by implementing the dynamic pricing strategy, with a recorded 15% rise in overall revenue within the first quarter of deployment.

Project Title: COVID-19 Risk Assessment Using Advanced Machine Learning Techniques

- Assessed COVID-19 patients' risk by utilizing various classification methods and resolving data imbalance with the SMOTE technique.
- Applied multiple machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, GBoost, and XGBoost, to create robust predictive models for assessing COVID-19 risk.
- Encapsulated the XGBoost model into an API using Flask, and successfully deployed the service on Google Cloud Platform (GCP), ensuring scalability and accessibility.

PIDILITE USA INNOVATION CENTER, Raleigh, NC 07/2017 – 11/2018

Data Scientist

- Used Design of Experiment (DOE) statistics tool to design experiments; collected, analyzed, and visualized experiment data.
- Coordinated tasks with other team members and reported project progress to business stakeholders.
- Improved product properties within 3 months by optimizing experiment conditions.

MICHIGAN STATE UNIVERSITY, East Lansing, MI 05/2016 – 06/2017

Data Scientist

- Built a regression model (R2 score was close to 0.98) to calculate absorption of can coating film to fragrance compounds from beverage, which helped Coca-Cola Company replace previous laborious method, saving tedious labor work and money.

LOUISIANA STATE UNIVERSITY, Baton Rouge, LA 08/2008 – 12/2010

Data Scientist

- Studied how temperature, humanity, and crop types affected CO2 content in soil and air, using statistic methods.

HENAN NORMAL UNIVERSITY, Xinxiang, Henan, China 06/2006 – 07/2008

Data Scientist / Research Associate

- Designed and conducted experiments and mentored undergraduates and master graduates in projects.

Education | Certifications

- **Doctor of Philosophy (PhD)**, Polymer Chemistry, University of Akron, Akron, OH
- **Master of Science (MS)**, Beijing University of Chemical Technology, Beijing, China
- **Bachelor of Science (BS)**, Henan Normal University, Xinxiang, China
- **Certified Professional Forecaster (CPF)**, Institute of Business Forecasting & Planning, online