Ruthvik Yelthuri

Data Analyst

yelthuriruthvik9@gmail.com +1 (940) 843-3207 Denton Texas TX76201 linkedin

Summary

As a **data analyst** with **2 years of experience,** I have honed my skills in leveraging data to drive business outcomes. I am proficient in using **SQL**, **Python, R**, and **data visualization tools** to analyze **complex datasets** and **extract meaningful insights**. My experience in construction documents analysis has equipped me with a strong understanding of **industry-specific data and processes**. I have successfully applied **advanced data analytics techniques** to address a variety of challenges, including customer **churn prediction**, **recommendation systems**, **image classification**, **NLP**, and **time series forecasting**. My ability to **communicate complex findings** effectively and **collaborate with cross-functional teams** makes me a valuable asset for any organization seeking **data-driven solutions**.

Skillset

- **Programming Languages & Frameworks**: Python, R, SQL, JAVA, HTML, CSS, NodeJS, AngularJS
- **Visualization Tools & IDEs**: Tableau, Power BI, Microsoft Excel, Visual Studio Code, PyCharm, Anaconda, Nmap, MARS
- **Packages**: NumPy, Pandas, Matplotlib, SciPy, ggplot2
- **Database**: MySQL, SQL Server (SSMS, SSIS, SSAS)
- **Cloud Platforms & ETL**: AWS (Redshift, EC2, S3, Glue, Lambda), Azure (Synapse Analytics, VMs, Data Factory, Database Migration Service), GCP (BigTable, Dataproc, Dataflow, Dataprep, BigQuery)
- **Methodologies**: SDLC, Agile and Waterfall

Professional Experience

Principal Advanced Analytics Jul 2024 – present | Dallas, USA

AT&T

Project: Customer Lifetime Value (CLTV) Optimization

- Built a sophisticated CLTV model using **survival analysis**, accurately predicting customer lifetime value with a mean absolute error of **5%**.
- Integrated data from **3 disparate sources**, including CRM, sales, and finance systems, using **Azure Data Factory** and **Azure Synapse Analytics**.
- Identified the top **20%** of high-value customers, accounting for **80%** of total revenue, using **clustering techniques** like **K-means**.
- Developed targeted marketing campaigns, including personalized email recommendations and loyalty programs, that increased customer engagement by **25%** and average order value by **15%**.
- Demonstrated a **ROI** of **3:1** on customer-centric initiatives**.**

Project: Customer Churn Prediction

- Developed a state-of-the-art customer churn prediction model using **XGBoost**, achieving an **AUC score of 0.92**, outperforming baseline models by **10%.**
- Leveraged **Databricks** to process over **10 million customer** records within 24 hours, optimizing data pipelines for efficiency through **Spark SQL** and **Delta Lake**.
- Engineered **25 key features**, including customer tenure, average monthly spend, and usage patterns, using feature engineering techniques like **one-hot encoding** and **normalization**.
- Deployed the model into a real-time production environment on **AWS Lambda**, enabling near-instantaneous predictions.
- Quantified the financial impact of churn reduction at **$1.5 million,** based on customer lifetime value analysis.

Construction Documents Analyst Jun 2021 – Jun 2022 | Anantasagar, India

NEBCO. INC

Project: Streamlined Bid Preparation Process

- Implemented a standardized workflow for document review and analysis, reducing bid preparation time by **20%**.
- Developed a **comprehensive checklist** for essential document requirements, ensuring completeness and accuracy**.** The checklist included items such as **project scope, drawings, specifications, permits, and site conditions**, ensuring that all necessary information was captured.
- Utilized **Bluebeam's markup tools** to efficiently highlight relevant scopes of work, **improving clarity** and **reducing errors.** By using Bluebeam's **advanced annotation features**, we were able to clearly mark specific areas of the documents that were relevant to **structural engineering**, reducing the risk of misunderstandings and errors.

Project: Enhanced Bid Accuracy

- Identified and **corrected errors** in **construction documents**, resulting in a **10% increase** in bid win rate.

- Developed a **quality assurance process** to review documents for consistency and completeness.
- Collaborated with estimating team to ensure accurate interpretation of **construction plans and specifications**. By working closely with the estimating team, we were able to **clarify any ambiguities or uncertainties** in the documents, ensuring that our bids were based on a clear understanding of the project requirements.

Construction Project Coordinator (Intern) Jan 2021 – Apr 2021 | Anantasagar, India
NEBCO. INC
Project: Site Inspection and Documentation
- Conducted thorough **site inspections** to assess project conditions, identify potential challenges, and collect relevant data.
- Prepared detailed **site reports**, including photographs, measurements, and observations, to document project progress.
- Assisted in the **development of site layout plans** and construction schedules.

Education

Master of Science in Advanced Data Analytics Aug 2022 – May 2024 | Denton, USA
University of North Texas

Bachelor of Eng. in Civil Engineering Aug 2017 – Apr 2021 | Anantasagar, India
SR University

Projects

Project: Analysis of Youth Tobacco Usage in the United States
Capstone Project
- Spearheaded an end-to-end machine learning project analyzing **Youth Tobacco Survey** dataset comprising **10,601 records** and **31 features** to predict **tobacco usage patterns** among US adolescents.
- Implemented robust **ETL pipeline** utilizing **pandas** for missing value imputation, **Z-score** based **outlier removal**, **feature engineering**, and **PCA-based dimensionality reduction** with **standardization.**
- Architected multiple models (**Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, SVR**) with comprehensive **cross-validation** and .**hyperparameter optimization** frameworks.

Project Outcomes
- Achieved **99.98%** accuracy (**R^2 score**) with **Linear Regression model**, demonstrating **RMSE** of **0.011195** through **systematic model tuning** and **validation protocols**.
- Engineered **statistical visualization** suite incorporating **box plots**, **learning curves**, and **comparative performance metrics** for **model evaluation** and **pattern analysis**.

Time Series Forecasting of Sales Data
Milestone Project
- Forecasted monthly sales for a retail company using a combination of **ARIMA** and **LSTM** models.
- Performed **Data preparation**, cleaned and preprocessed the sales data, handling missing values and outliers.
- Stationarized the **time series data** using **differencing and trend-seasonality decomposition**.
- **Optimized hyperparameters** for ARIMA and LSTM models using **grid search**.
- Evaluated model performance using metrics with mean squared error (**MSE**), mean absolute error (**MAE**), and root mean squared error (**RMSE**).
- **Deployed the forecasting model** to provide **insights**.

Project Outcomes
- Achieved a **mean squared error** (MSE) of **0.025** on the **validation set**, demonstrating **accurate forecasting performance**.
- Identified **seasonal patterns** and **trends** in **sales data**, providing **valuable insights** for **demand planning**.
- Accurately predicted **peak sales periods** and **fluctuations** due to external factors (e.g., holidays, promotions).

Sentiment Analysis of Customer Reviews
Milestone Project
- Developed a **sentiment analysis model** using a **bidirectional LSTM network** with attention mechanism.
- Utilized word embeddings (**Word2Vec, GloVe**) to capture **semantic relationships between words**.
- Utlized techniques like **attention visualization** to understand which **words or phrases contribute** most to the **predicted sentiment**.

Project Outcomes
- Accurately classified **customer reviews into positive, negative, and neutral sentiments**.
- **Identified key themes** and **topics** within **customer reviews**, providing valuable insights for product development.
- Achieved an accuracy of **88%** on a **sentiment classification task** with a **labeled dataset of customer reviews**.