**Blessy Thomas**

West Hempstead, New York 11552

(516) 884-0331

[bthomasca24@gmail.com](mailto:bthomasca24@gmail.com)

<https://github.com/bthoma16/IST-782-Applied-Data-Science-Portfolio>

<https://www.linkedin.com/in/blessy-thomas-bthomasca24/>

**Summary**

As a recent Master of Science in Applied Data Science graduate from Syracuse University, I bring a robust analytical background with hands-on experience in data analysis, machine learning, and database management. Specializing in predictive modeling and NLP, I have developed high-accuracy models and streamlined data processes, as evidenced by a predictive attrition model with 82.65% accuracy and a text classification model surpassing baseline accuracy by 4%. With over two years of professional experience in data-driven roles, I have enhanced operational efficiencies and contributed to the integrity of laboratory results. My collaborative approach, combined with strong project management and communication skills, positions me to deliver impactful data insights and drive business decisions. Recognized as a coauthor of a published GenBank annotation, I am adept at translating complex data into actionable strategies, ready to excel as a Data Analyst.

**Education**

**Syracuse University** – Master of Science in Applied Data Science May 2024

**State University of New York at Old Westbury** – Bachelor of Science in Biological Sciences May 2021

**Skills and Tools**

* Regression analysis, Time series analysis, Machine Learning, Natural Language Processing
* Python, R, SQL
* Microsoft SQL Server, MongoDB, MySQL, Laboratory Information Management System
* Project management, Impact evaluation, Microsoft Excel, ShareFile, Asana, Canva
* MAXQDA, Audio Transcription, Google Analytics
* Azure Data Studio, RStudio, Jupyter Notebook

**Data Science Projects**

* **Analyzing Company Attrition:** Led a comprehensive analysis on employee attrition using a Kaggle HR analytics dataset. Utilized Python in Jupyter Notebook and Google Colab for data cleaning and preprocessing, tackling duplicates and missing values. Conducted exploratory data analysis to uncover key retention factors such as age, travel frequency, and performance. Employed pandas, matplotlib.pyplot, and numpy for data manipulation and visualization. Findings highlighted higher attrition among younger employees and frequent travelers, while top performers showed strong retention. Developed a predictive model with 82.65% accuracy to identify at-risk employees, resulting in actionable recommendations to reduce turnover, focusing on support for younger staff and travel policy adjustments.
* **Classification of Text:** Developed a sophisticated text classification model to discern sentiments in movie reviews, leveraging Python and Sci-Kit Learn. Managed a dataset of 156,060 phrases from a Kaggle competition, employing advanced preprocessing and feature selection methods such as bag of words and bigrams. Extensive testing across various classifiers, including Naïve Bayes and logistic regression, led to a notable enhancement in classification accuracy by optimizing feature sets. Results highlighted the critical role of preprocessing in boosting model accuracy. Notably, Logistic Regression and Linear SVC with the SL + LIWC feature set achieved the highest performance, reaching 59% accuracy, surpassing the unigram baseline of 55%.
* **Inventory Management Database:** Developed and led the transition of Whole-ER Foods' inventory management from Microsoft Excel to a comprehensive database system utilizing Azure Data Studio and Microsoft SQL Server. This project significantly improved inventory tracking, product identification, and operational efficiency by introducing a robust relational database. The initiative streamlined inventory processes and provided hands-on experience in database design, development, and data analysis, showcasing the application of theoretical knowledge to real-world challenges.
* **Student Grade Prediction:** Forecasted academic risks among high school students using data mining. Processed a Kaggle dataset with 395 records and developed predictive models including decision trees, logistic regression, k-NN, SVM, and random forests. The final random forest model yielded a 99.28% accuracy rate in pinpointing at-risk students, offering valuable data for preemptive educational support. Key factors impacting final grades were identified: failures, freetime, fedu, dalc, walc, schoolsup, absences, age, famrel, and famsize.

**Professional Experience**

**Passion for Life, Inc.**  Atlanta, GA (Remote)

**Research and Evaluation Intern** June 2023 - Present

* Assisted in the design and implementation of data collection instruments, such as pre/post surveys and focus group protocols, to evaluate program participants, ensuring high-quality data collection for accurate program assessment.
* Developed and reviewed surveys, forms, and other data collection activities, optimizing content and format to promote clarity and responsiveness, thereby improving participant engagement and data quality.
* Identified program needs through meticulous data collection and analysis, providing critical insights that informed program adjustments and enhancements to better meet participant requirements.
* Designed research presentations and reports for key stakeholders, effectively communicating program outcomes and securing ongoing partnerships and funding by demonstrating the value and impact of the program.

**Acūtis Diagnostics**  Hicksville, NY (On-site)

**QC Specialist**  January 2022 – Present

* Conducted comprehensive quality checks on toxicology accessions, ensuring accurate test orders and patient information alignment with requisition forms, thereby preventing costly billing errors and unnecessary testing.
* Analyzed medical data within CGM LABDAQ, contributing to the integrity and reliability of laboratory results and downstream billing processes.
* Resolved toxicology holds in ShareFile by updating orders with correct patient information and tests as per client resolution forms, effectively saving costs and streamlining billing operations.
* Reviewed and rectified integration errors, ensuring the successful incorporation of requisitions into CGM LABDAQ, which facilitated accurate billing and reduced potential revenue loss.

**Specimen Accessioner** August 2021 – Present

* Accurately accessioned approximately 1,000 diverse specimens daily, including COVID-19 and toxicology samples, ensuring timely delivery to clients and contributing to the laboratory's reputation for reliability.
* Enhanced data integrity by meticulously verifying patient information and resolving discrepancies, leading to a decrease in data errors and an increase in client trust and satisfaction.
* Played a pivotal role in public health by processing critical COVID-19 tests, aiding in the containment efforts of the pandemic and demonstrating adaptability in high-stakes environments.
* Innovated problem-resolution protocols for specimen types and missing information, which are now standard practice, enhancing the laboratory's problem-solving efficiency and reducing specimen rejection rates.

**Healing Touch Physical Therapy & Rehabilitation P.C.**  New Hyde Park, NY (On-site)

**Physical Therapy Aide and Receptionist**  June 2021 - August 2021

* Managed the application of therapeutic modalities, such as cervical and lumbar heat/ice packs, for an average of 30 patients per day, ensuring optimal patient comfort and aiding in recovery.
* Assisted an average of 90 patients weekly with mobility, contributing to their therapy progress and enhancing patient satisfaction.
* Enhanced patient communication by proactively scheduling follow-ups and appointment reminders, leading to a 90% patient attendance rate.
* Successfully managed sensitive patient information and processed payments, maintaining a 100% compliance rate with privacy regulations and financial accuracy.

**ParentChild+**  Mineola, NY (On-site)

**Research and Data Intern**  June 2019 - August 2019

* Spearheaded the transcription of over 50 hours of audio files from interviews and focus groups, ensuring accurate data capture for qualitative analysis.
* Utilized MAXQDA software to perform thematic and content analysis, enhancing the depth of insights for program evaluation reports.
* Compiled and presented descriptive statistics for internal reports, influencing strategic decision-making by providing data-driven recommendations.
* Implemented rigorous data quality checks and cleaning protocols on large datasets, improving data accuracy by 20% and ensuring reliability of research findings.

**State University of New York at Old Westbury – Biology Department** Old Westbury, NY (On-site)

**Phage Hunting Research Assistant**  September 2018 – December 2018

* Successfully isolated and characterized multiple Actinobacteriophages specific to Mycobacterium smegmatis, contributing to the expansion of the phage database used for research and educational purposes.
* Annotated the genomes of isolated bacteriophages with precision, achieving a 98% accuracy rate in genomic identification and mapping, thereby enhancing the understanding of phage genetics within the scientific community.
* Leveraged bioinformatics tools and software to analyze phage DNA sequences, becoming the go-to expert for data interpretation and troubleshooting within the research group.

**Achievements**

**Mycobacterium phage MilanaBonita, complete genome**  Old Westbury, NY (Remote)

**GenBank** September 8th, 2021

* Coauthor of the published GenBank annotation of the phage genome, MilanaBonita.