Bryan **Mejia**

Queens, New York | m: +1 347 791 3404 | bryanmejiapersonal@gmail.com | LinkedIn | Website

Data Scientist | Software Engineer

Career Profile

Proven talent in full-stack development and applied data science, designing and implementing robust applications and predictive models that solve real-world problems. Experienced in leveraging Python and libraries like Scikit-learn within the Anaconda environment to engineer features, build, evaluate and optimize predictive models for complex datasets, with a focus on delivering actionable insights. A results-driven professional with a strong foundation in teamwork, communication, and organizational skills,consistently collaborating with teams to deliver high-quality, documented software.

Core Competencies

- Software Development
- System Design
- Debugging and Testing
- Object-Oriented Programming
- Problem Solving & Analysis
- Project Management
- Task Prioritization
- Technical Communication
- Git/Github
- Data Analysis (EDA)
- Full-Stack Development
- Predictive Modeling

Technical Experience

HEARTSCOPE FULLSTACK APP

- Engineered a 3-layer architecture (UI, Algorithms, Data) using Java (backend) and JavaSpring + JavaScript/HTML/CSS (frontend), ensuring modularity and scalability.
- Optimized preprocessing pipeline: handled missing values with modal imputation, applied one-hot encoding for multi-class features, and binary encoding for categorical features.
- Achieved ~88% prediction accuracy on testing dataset (error rate ~12%) against a target of ≤10%.
- Developed and implemented a Random Forest algorithm, empowering users with personalized, data-driven insights for proactive health management.
- Optimized performance: prediction pipeline executed in milliseconds per user input, enabling real-time feedback (well under the 10-second requirement).
- Ensured project quality and maintainability through the creation of 35 pages of IEEE-compliant software documentation, demonstrating a commitment to professional standards.

MENTAL HEALTH DATA ANALYSIS & PREDICTIVE MODELING

- Engineered a robust predictive model for mental health data by developing and rigorously evaluating over 7 classification models (e.g.,Random Forest, SVM, Logistic Regression) using imbalance handling strategies.
- Achieved a 96% balanced accuracy on a Decision Tree model and a consistent 83% average balanced accuracy on a Random Forest classifier through 5-fold cross-validation, demonstrating effective handling of a class-imbalanced dataset.
- Analyzed ~2,000+ survey records containing demographic, lifestyle, and academic/work factors to predict depression risk.
- Addressed class imbalance using balanced accuracy metrics and 5-fold cross-validation, improving reliability by 15–20% compared to accuracy alone.

- Optimized model performance through hyperparameter tuning and comprehensive evaluation using key metrics such as Precision, Accuracy, Recall, Specificity, and F1-Score.
- 5+ insightful graphics, providing a deeper understanding of the dataset and directly guiding model development

SPAM EMAIL CLASSIFICATION ANALYSIS

- Engineered a comprehensive machine learning pipeline to classify emails as spam or not spam, handling a dataset of 5,172 email records and 3,001 features.
- Conducted a detailed class imbalance analysis, identifying that spam emails accounted for 29% of the dataset, and implemented stratified sampling and class weighting to address the imbalance.
- Trained and evaluated 6 distinct classification models, including Random Forest and Logistic Regression, on a 80/20 train-test split to ensure robust performance evaluation.
- Achieved a peak accuracy of 99% with a Stacking Classifier and a 97% accuracy with a Random Forest Classifier, demonstrating exceptional model performance.
- Minimized false negatives and false positives, with the Stacking Classifier achieving a 99% recall and 97% precision for the spam class, making it ideal for a real-world spam detection system.

Education and Credentials

Bachelor of Arts (B.A.) in Computer science –Queens College 2025;

Additional Information

Languages: English, Spanish

Technical Proficiencies: Windows, Linux (PopOS, Ubuntu), Mac OS, Microsoft Office , servers, routers, wireless LAN, C++, Java, Python, Javascript (Express, Node, React), Anaconda Environment, Git/Github, HTML, CSS, JavaScript, Python, SQL, Slack

Interests: Hiking, Reading, Social Media, Music, Weight Training, Mixed Martial Arts, Cycling, Traveling