

Surya Pranav Sukumaran

10 Trapelo St – Brighton – MA 02135
☎ +1 503-560-1009 • ✉ pranavs@bu.edu

EDUCATION

Boston University

MSc Applied Data Analytics

Key Modules: Data Mining, Database Design and Implementation, Data Science using Python, Advanced Machine Learning, Big Data Analytics

USA

09/2022–Present

Asian Institute of Technology

MSc Business Analytics and Digital Transformation

Key Modules: Leadership & Organizational Management, Change Management, Storytelling Using Data

Thailand

08/2020–12/2021

University of Exeter

BSc Medical Sciences

UK

09/2016–09/2020

SKILLS

Programming Languages: Python, R, SQL, HTML, CSS

Packages: Pandas, NumPy, Dplyr, Matplotlib, Seaborn, ggplot2, Plotly, Scikit, Folium

Tools: Apache Spark, Hadoop, Google Cloud Services, AWS, Docker, Apache Zeppelin, Tableau, Jupyter Notebooks, Google Colab, Anaconda, GitHub

PROFESSIONAL EXPERIENCE

Brands On Road Media Co.

Data Analyst Intern

Thailand

05/2022–08/2022

- Extracted GPS data from MongoDB database to illustrate taxi activity in Bangkok through heatmaps, offering a detailed view of advertisement visibility zones for clients.
- Contributed on client reports containing advertisement performance metrics, utilizing Tableau and Python (Plotly) for visualizations sourced from database data via SQL queries.
- Reports guided clients in strategically optimizing their taxi-based advertising campaigns, providing actionable insights through heat maps for location hotspots, and time-series plots to identify peak activity times.

Kids Kreations Co.Ltd.

Marketing and Data Analyst for Online Sales

Thailand

08/2021–03/2022

- Contributed to improving the company's data infrastructure by implementing a PostgreSQL database, optimizing tables, constraints, and stored procedures, resulting in a significant improvement in query performance and efficient management of product, shipping, and customer data.
- Designed an Entity-Relationship diagram and mapped it to a relational schema, ensured tables were normalized and keys were optimized for efficient querying.
- Extracted marketing insights by writing and executing SQL queries and visualized best selling items by category, location insights, and customer purchasing trends using Matplotlib, helping the marketing team to target their campaigns more effectively.

PROJECTS

Visualization of Big Data: Apache Zeppelin

9/2023–10/2023

- Explored visualization of big data by leveraging Apache Zeppelin for seamless integration with Amazon EMR clusters, facilitating the processing and analysis of expansive NYC Taxi Trip datasets.
- Implemented dynamic forms and SQL within Apache Zeppelin notebooks to query and manipulate vast datasets swiftly, enabling rapid visualization and immediate data interaction; this approach significantly reduced the time-to-insight for large-scale data.
- Utilized Folium, a Python library, to transform preprocessed and aggregated big data into detailed geospatial visualizations, creating interactive maps that highlight taxi traffic density, routes, and hotspots.

Mental Health Insights Using Classification Algorithms

4/2023–4/2023

- Performed a machine learning classification analysis using Python on public mental health data reported within the US criminal justice system, highlighting intersections between trauma disorder and demographics.
- Trained and compared multiple machine learning models including Logistic Regression, Naive Bayes, SVM, and Random Forest, with the Naive Bayesian model exhibiting superior recall and F1 scores.
- Assessed coefficient weights across models to identify age, 65 and older, as a significant feature correlating with traumatic stress disorder.

Counterfeit Banknote Detection

2/2023–3/2023

- Conducted a regression statistical analysis on a dataset from UCI's Machine Learning Repository containing 1372 samples and 5 attributes, utilizing image features derived from wavelet transformation to detect counterfeit notes.
- Trained and developed a logistic regression model, achieving a 90.8% accuracy on the training set and 90.5% on the testing set. Despite challenges like class imbalance and potential overfitting, parameter tuning and data preprocessing helped in maintaining model strength.
- Utilized odds ratio to interpret the influence of each feature on counterfeit detection, with cross-validation results affirming the model's strength.
- Identified 'Entropy of the image' (V4) as a strong predictor for counterfeit detection while controlling for other features, examined model assumptions and addressed multicollinearity by excluding some features to ensure model validity.

Keyword Extraction Using BERT

2/2023–3/2023

- Investigated keyword extraction from political articles in VOX (left-leaning) and Talking Points Memo (right-leaning) using a pre-trained BERT model; compared original and summarized text for keyword consistency.
- Utilized KeyBERT for keyword extraction and calculated cosine similarity to measure consistency between original and summarized article keywords, ensuring the preservation of key information and context.
- Enhanced keyword extraction precision by adjusting the n-gram range from 2 to 3, observing improvements in cosine similarity scores for keywords in summarized VOX articles compared to Talking Points Memo articles.