

Shradhit Subudhi

shraddhit.com | linkedin.com/in/shraddhit/ | github.com/shraddhit

Professional Experience

ZS Associates, Inc. | New York City | Data Scientist | Feb 2020 - Present

Designed, developed, and executed high-impact analytics solutions, leveraging expertise in handling large volumes of structured and unstructured data. Utilized data analysis techniques to empower clients in making informed decisions based on factual insights.

Omnichannel Dynamic Targeting

- Led a team of 2 Data Scientists and 2 Data Engineers to support a big pharma client
- Enhanced, Operationalized and, Maintained dynamic targeting pipeline
- Delivered a weekly call plan for Neuroscience and Infectious Diseases sales representatives, incorporating target customers and personalized prescribing messages for effective customer engagement
- Achieved a significant 35% improvement in sales through the implementation of dynamic targeting strategies

Computational Phenotyping

- Spearheaded the implementation of an AI tool to generate a generalizable vector representation of a patient's medical state
- Applied NLP and DL techniques to obtain medical code-level vector embeddings
- Employed probability-based weight averaging techniques to convert medical code vectors to the patient level
- Constructed a patient similarity network to maximize homogeneity within and heterogeneity among groups and accurately identified spatial and temporal patterns that distinguished diseases (e.g., NSCLC, AD, CKD, MS)
- Demonstrated a remarkable rand index of 70% for disease identification
- Published a white paper on the project, accessible at:

Synthetic Control Arm

- Led the development and implementation of techniques for conducting clinical trials with a synthetic control arm
- Identified similar patients in real-world data using inclusion/exclusion criteria
- Utilized propensity score matching methods to minimize bias when comparing treatment and control datasets
- Calculated treatment effects for randomized control trials

Covid Visitation Monitor

- Developed a comprehensive tool to monitor visitation policies across hospitals and group practices in the US.
- The output of the tool played a pivotal role in amending the call plan for the sales representative team, aligning with access rules and regulations for big pharma sales representatives.

Veritonic, Inc. | New York City | Data Scientist, Intern | June 2019 - Dec 2019

Developed deep neural network data science modules for precise gender, emotion, and instrument detection in audio data. Enhanced comprehension of sound resonance in radio and podcast advertisements for target audiences. Leveraged automated tags to rate audio, augmenting ROI for products and client firms. Key Aspects of the project below:

- Applied audio analysis techniques such as MFCC extraction and spectrogram generation to extract meaningful features from the audio data.
- Developed and implemented deep learning models to enhance the selected metrics for the project. Deployed the models effectively, ensuring successful integration & utilization.
- Utilized tools and frameworks like Librosa, Keras, TensorFlow, and Sklearn to support the project's implementation and analysis processes.
- Demonstrated expertise in audio analysis, deep learning, and model deployment through successful project execution.

Education

Rutgers University, Rutgers Business School, NJ | Sep 2018 - Dec 2019

Masters in Information Technology and Analytics | GPA: 3.717

College of Engineering Pune, India | May 2014 - June 2018

Bachelor of Technology in Computer Engineering | SGPA: 8.62

New Jersey 07029

(609) 325-6855

shraddhit@gmail.com

Technical Summary

Languages: Python, R, SQL, C/C++, git, HTML/CSS, PySpark

Libraries/Frameworks: Scikit-Learn, Pandas, Numpy, TensorFlow, PyTorch, Plotly, GraphX
Tools: Excel, Word, PowerPoint, Access, G Suite, Jekyll, Power BI, Tableau, Dataiku

Knowledge Area: Supervised, Unsupervised, regression, pattern, trend identification, TensorFlow, Anomaly detection, time series, outlier detection, data mining, visualization, deep learning, computer vision, business intelligence, redshift, recommendation engines, Reinforcement learning, data warehousing, MongoDB, document clustering, text classification, Spark RedShift

Certifications

Introduction to Data Science in Python by University of Michigan

Applied Plotting, Charting & Data

Representation in Python by University of Michigan

Applied Machine Learning in Python by University of Michigan

Applied Text Mining by University of Michigan

Leadership Roles

Founder & Co-Secretary, Software Development Section, CoEP

Team Lead, Smart India Hackathon

Class Representative, B.Tech Computer Engineering from 2014-2018

Web Team Head for ED Cell, COEP

Mathematics Teacher (Volunteer) at St. Jude India

Achievements

7th Rank in Public Computer Vision Challenge, identifying covid patients by analyzing the CT scan image

4th in B.Tech Project Competition for Anecdote Detection System held at College of Engineering Pune

Among **Top 1 percentile** in Grade 12th HSC - Maharashtra, India State Board

Elaborate Projects Explanations

[ZS] Omnichannel Dynamic Targeting: Led a team of 2 Data Scientists and 2 Data Engineers to develop and operationalize dynamic targeting algorithms using advanced AI and optimization techniques for Neuro Science and Infectious diseases, in collaboration with a major pharmaceutical client.

The project aimed to enhance the effectiveness of customer engagement by delivering quarterly and weekly call plans that included target customers and prescribing messages.

To achieve this, we implemented a comprehensive data pipeline that ingested data from diverse sources such as IQVIA (Retail Sale, Prescriber Dynamics), LAAD Claims, FIA Claims, SHS Claims, DDD (non-retail sales), MMIT (formulary & lives), Speciality pharmacy (prometric), and Veeva pulse.

The project encompassed several interconnected components:

- **HCP Universe:** We filtered and identified relevant healthcare professionals based on their therapeutic area and prescribing behavior, collecting necessary features.
- **Target Augmentation:** Working closely with the pharma brand team, we refined the HCP targeting criteria to align with their preferences and objectives.
- **Feature Engineering Pipeline:** Leveraging a master dataset, we developed an input pipeline for the Explainable AI (XAI) models. The dataset consisted of HCP and date-week level information, utilizing 105 weeks (2 years) of historical data to predict future NBRx sales.
- **Explainable AI (XAI):** Using regression models, we predicted future HCP values for writer HCPs (based on NBRx sales). For non-writer HCPs, we employed classification models to estimate the probability of becoming writers in the future. HCP value was determined based on the predicted NBRx sales. Through SHAP analysis, we gained insights into the output of the underlying models, providing explanations for targeting specific HCPs and understanding the reasons behind their selection.
- **Scenario Generation:** Utilizing the predictions from the XAI models, we generated a scenario generation table that encompassed various possible scenarios. This allowed us to explore changes in call counts and predicted HCP values while keeping other features constant.
- **Channel Optimization (CO):** Using convex optimization techniques, we optimized the number of calls for each physician, taking into account constraints such as total call/email limits per representative and trending logic.
- **Downstream Integration:** The output of the Channel Optimization phase was seamlessly integrated into c360, a platform that provided a user-friendly interface for sales representatives to access the optimized call plans.

In addition to the core pipeline components, we incorporated other functionalities:

- **Feedback loop logic** allowed sales reps to tag Dead, Moved, or Retired HCPs, ensuring they were excluded from the target list.
- A message optimization pipeline generated customized email suggestions based on target HCPs' behavior, further enhancing engagement.

- Trending logic constrained short-term outputs to align with quarterly outputs from the long-term pipeline.

To facilitate the development, deployment, and management of AI applications, we leveraged Dataiku, an enterprise AI platform. In the backend, we utilized AWS EMR and EKS technologies. The majority of the code was written in Python, PySpark, and SQL, with certain components implemented in R and Spark R.

Overall, the implementation of the Omnichannel Dynamic Targeting initiative resulted in a significant 35% improvement in sales, demonstrating the effectiveness of the developed algorithms and strategies.

[ZS] Covariate Adjustment: As an experienced clinical RnD subject matter expert contributed to the development of a Point of View on Covariate Adjustment for ZS's business development. Demonstrated expertise in providing insights and guidance on implementing covariate adjustment methods to enhance and reduce bias during the clinical trials.

[ZS] Quality Risk Management:

- Assisted in the business development efforts for quality risk management services specific to clinical trials, targeting pharmaceutical and biotechnology companies.
- Collaborated with cross-functional teams, including data scientists, clinical research experts, and regulatory specialists, to develop comprehensive risk management strategies for clinical trials.
- Lead client meetings and presentations, effectively articulating the value proposition of quality risk management in clinical trials and highlighting the firm's expertise from Data Science / Automation perspective. Understand the clients requirements of risk-based approaches in their clinical trial operations, ensuring compliance with regulatory requirements and enhancing overall trial quality, and trying to pitch in up-to-date AI developments to automate their processes.

[ZS] Webinar Attendance Optimization

Led a team of 2 data scientists to improve the webinar attendance for the Bexsero Meningitis vaccine. Clients requested our expertise in utilizing AI/ML techniques to identify effective email types and phrases for engaging healthcare professionals (HCPs). However, we discovered that the clients lacked diversity in their email targeting strategy.

We assisted in identifying similar target audiences and provided strategic guidance, emphasizing the importance of personalizing email content. This enabled the clients to enhance their future email marketing efforts by tailoring messages to specific recipients and improving engagement.

[ZS] Driver's Analysis: Developed a predictive model to help understand the drivers to build a launch strategy for novel obesity drugs in the market based on market survey. Build a UI tool on Jupiter notebook, with ML model in the backend & generated SHAP reports.

[ZS] CT Scan Covid Classification Public Competition

7th Rank in Public ML Challenge, predicting if a person has COVID looking at Lung CT Scan Images

- CT scans can be used to detect COVID-19, but the process of manually analyzing these images is

time-consuming and labor-intensive. This can lead to delays in diagnosis and treatment, which can have serious consequences for patients.

- Developed a deep learning model that can classify CT scan images of COVID-19 patients with high accuracy. This model can be used to automate the process of image analysis, which can free up radiologists to focus on other tasks.
- My model achieved an accuracy of 98.91% on a test set of CT scan images. This is significantly higher than the accuracy of human radiologists, who typically achieve an accuracy of around 85%.
- The model can help to improve the diagnosis and treatment of COVID-19 patients. It can also help to reduce the workload of radiologists, freeing them up to focus on other tasks.

[Veritonic] Genre Detection : A Music Information Retrieval project i.e. classifying audio music files into various popular genres. Built spectrogram & extracted MFCC from the .wav files (using FMA Dataset) using various CNN-RNN-based models & got an accuracy close to 70%.

[Choose To Think] Anecdote Detection System - B.Tech - Research Project - Pioneering application to automatically identify anecdotes analyzing a set of sentences as input using articles from Blinklist and contemporary techniques of natural language processing and Machine Learning. Key Aspects of the project include research, pattern and trend identification, statistics, rule-based estimation, natural language processing, Stanford NLP library, and Ensemble Models.

[Grad] Airbnb / Zillow - Kaggle Competition - Predicted in which country a new user will make his (or her) first booking. Tasks involved: data cleaning, exploratory analysis, and performance-boosting tactics. Zillow; predicted the important features that contribute in predicting the Z - Score and also ended up predicting the Z score using various ML models.

[Grad] Wine Quality Predictor : Analyzed the quality of Portuguese "Vinho Verde" wine given various physicochemical inputs. Tasks involved: Data Cleaning, Preprocessing, analysis of the dataset, visualization, and proficient R programming, fitted the model keeping in mind the assumptions of the models used. Reported an accuracy of more than 75% using Binomial Logistic Regression. Link: <https://rpubs.com/shradhit/winequality>

[UnderGrad] Sentence Completion - Research Project - To automate the task of Sentence Completion in SAT that involves choosing the correct word from the given set of word options such that the resulting sentence is semantically correct. Our team tried to automate the manual task using Natural Language Processing.

[UnderGrad] Brain Injury Score Calculator :The software takes the input of the patient's condition and outputs the severity of his injury using the Glasgow Scale using complicated mathematical calculations. The project was made on python, and the framework used was panda3d.

[UnderGrad] Repository for Publication Database - A web-based central database for storing details about all the papers published by the College of Engineering Pune & faculty with the hierarchy of access. The technology stack used - is LDAP, HTML, JavaScript, CSS, etc.