**KRUSHIKA TAPEDIA**

[tapediak29@gmail.com](mailto:tapediak29@gmail.com) | (774) 502 3298 | <https://www.linkedin.com/in/krushika29/>

**PROFESSIONAL SUMMARY:**

* With over **8 years** of experience in the field, I have worked on various artificial intelligence projects ranging from building data pipelines and designing legacy systems to implementing cutting-edge deep learning techniques for natural language processing and image analysis.
* Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile/ Scrum Methodologies.
* A highly experienced and accomplished data scientist with a proven track record of delivering exceptional results through data-driven approaches.
* Strong background in Statistics, Artificial Intelligence, and Software Development, which has enabled me to develop innovative solutions that have resulted in significant business impact.
* Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python, Power BI and Tableau.
* Hands-on experience on Python, R, and libraries like NumPy, Pandas, Matplotlib, Seaborn, NLTK, and Sci-Kit learn, SciPy, Jupyter, Pycharm, and Anaconda.
* Adept in implementing regression analysis, ANOVA, Hypothesis testing, deep learning, neural networks, and advanced ML algorithms such as Random Forest, XGBoost, Gradient Boosting Machine, Linear & Logistic Regression
* Experience in integrating data, profiling, validating and data cleansing transformation and data visualization using R and Python.
* Hands-on experience in implementing LDA, and Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis, and good knowledge on Recommended Systems.
* Proficient in Statistical Modeling and Machine Learning techniques (Linear, Logistics, Decision Trees, Random Forest, SVM, K-Nearest Neighbors, Bayesian) in Forecasting/ Predictive Analytics, Segmentation methodologies, Regression-based models, Hypothesis testing, Factor analysis/ PCA, and Ensembles.
* Theoretical foundations and practical hands-on projects related to (i) supervised learning (MLP, CNN, linear and logistic regression, boosted decision trees, Support Vector Machines, neural networks, NLP, BERT Model), (ii) unsupervised learning (clustering, dimensionality reduction (PCA and LDA), Time Series Analysis and Forecasting using ARIMA model), (iii) probability & statistics, experiment analysis, confidence intervals, A/B testing.
* Experience in designing star schema, Snowflake schema for Data Warehouse, ODS architecture.
* Expertise in TensorFlow and Keras to do machine learning/deep learning package in python.
* Highly skilled in using visualization tools like Tableau, ggplot2, and Power BI for creating dashboards.
* I am a recipient of several awards, including the Best Idea Award and First Place in a poster presentation. With my diverse technical skillset, passion for innovation, and dedication to delivering high-quality results, I am confident in my ability to make a valuable contribution to any data science team.

**EDUCATION**

* **Worcester Polytechnic Institute (WPI)**, Worcester, MA - 2019
  + Master of Science in **Data Science**, GPA - **4.0/4.0**
* **Savitribai Phule Pune University (SPPU), Pune, India – 2016.**
  + Bachelor of Engineering in **Computer Engineering**, GPA - **3.7/4.0**

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Languages** | Python, R, Java, JavaScript, HTML, CSS, C++, MATLAB |
| **Databases & Frameworks** | SQL, NoSQL, Hadoop, Spark, AWS, GCP, MS Azure, Airflow, Kafka, Graphite |
| **Platforms & Frameworks** | Django, Flask, IBM SPSS, Tableau, MS SSRS, Power BI, Dash, JMP, Databricks |
| **Data Science Packages** | TensorFlow, PyTorch, NLTK, Keras, Pandas, NumPy, SciPy, Theano, Scikit- learn, Seaborn, Matplotlib, Plotly, dplyr, ggplot 2 |
| **Machine Learning Techniques** | Linear, Logistics, Decision Trees, Random Forest, SVM, K-Nearest Neighbors, Bayesian |
| **Software Version Control & Documentation** | Git, Gerrit, JIRA, Confluence, Crucible, Latex |

**PROFESSIONAL EXPERIENCE:**

**Client: Abbott Rapid Diagnostics, San Diego, CA**

**Role: Senior Data Scientist Mar 2020 – Present**

**Abbott is a global healthcare leader that helps people live more fully at all stages of life. Its portfolio of life-changing technologies spans the spectrum of healthcare, with leading businesses and products in Diagnostics, Medical Devices, Nutritional, and Branded Generic Medicines.**

**Responsibilities:**

* Developed a machine learning model to predict equipment maintenance needs and recalls for eReader, a drug screening system, resulting in a 20% reduction in unplanned recalls and a 15% increase in equipment efficiency. Utilized historical data on equipment performance and maintenance logs to train the model, incorporating feature engineering and data normalization techniques to improve accuracy.
* Utilized Python's Pandas, NumPy, and Scikit-learn libraries to perform data cleaning, feature scaling, and feature selection to devise an optimized and efficient model.
* Implemented the model in a production environment via a web-based dashboard using Flask and Plotly in AWS, providing real-time notifications and visualization of predicted maintenance needs to maintenance personnel. Collaborated with cross-functional teams, including data engineers and IT personnel, to ensure successful deployment and adoption of the model.
* As an SME, re-designed legacy systems and built data pipelines in Apache Airflow to amalgamate data from various sources, leading to centralized cloud storage in Azure for easy data retrieval and analysis.
* Trained and integrated a sequential neural networks model within the database to automate assigning 3 out of 5 customer complaint categories, saving up to 3.5 hours of manual inspection per day for each technician.
* Developed and deployed a web application in production using Django (web framework in Python) to generate custom analysis reports for users of the QA department to process customer complaints faster, thereby eliminating a backlog of about 20,000 complaints within a week using an XGBoost model in AWS Sagemaker.
* Devised an image analyzer for the BinaxNOW COVID-19 Antigen Self-Test kit to determine the self-test result using the VGG-16 model via transfer learning technique in neural networks.
* Performed pre-processing and deep learning techniques like data augmentation, image segmentation, normalization, transfer learning, etc., on the self-test card images, achieving an f1-score of 99.7% using TensorFlow in Python.
* Conducted NLP-based projects, including sentiment analysis on customer complaints, utilizing natural language processing techniques and leveraging libraries such as NLTK and spaCy to extract insights and improve customer support processes.
* Executed Generative AI projects, developing models for generating synthetic data to augment the training set, enhancing the robustness and performance of machine learning models.
* Collaborated on AI projects in both Microsoft Azure and AWS platforms, leveraging services such as Azure Machine Learning, AWS Sagemaker, and Databricks for various tasks such as model training, deployment, and data processing.
* Performed and conducted several code review activities as an SME, adhering to FDA validation standards and ensuring software quality and compliance with regulatory requirements.

**Environment**: MS Azure, Python, NumPy, Pandas, Scikit Learn, Flask, Databricks, AWS, Apache Airflow, Django, XGBoost, AWS Sagemaker, TensorFlow, NLTK, spaCy.

**Amadeus North America Inc., Waltham, MA**

**Role: Data Scientist Jan 2018 – Feb 2020**

**Amadeus’ solutions connect travelers to the journeys they want, linking them via travel agents, search engines and tour operators to airlines, airports, hotels, cars and railways. Amadeus’ Technology has always been critical to developing Global Travel, Increasing Scale, Choice and Access.**

**Responsibilities:**

* Conducted research and analysis of more than 42 factors affecting business travel stress and performed feature engineering and factor analysis to identify the most significant factors for stress prediction.
* Developed and implemented a hybrid model architecture using multiple artificial intelligence or machine learning techniques such as classification using ensemble methods like random forests and logistic regression to predict a business trip's and a traveler's stress levels based on approximately 5 million flight records.
* Designed and developed services in Python and R to predict the stress level of a trip and a traveler, achieving a 77% accuracy rate.
* Integrated the stress prediction module into the Amadeus Booking Tool architecture and provided recommendations for alternative travel suggestions when high stress was identified.
* Collaborated with the Amadeus development team to ensure seamless integration of the stress prediction module into the booking tool, enhancing user experience and improving the overall quality of service.
* Conducted extensive testing and validation of the predictive model and ensured data privacy and security regulations compliance.
* Maintained and optimized the predictive model to ensure its accuracy and reliability over time.
* Led end-to-end software development lifecycle (SDLC) processes, driving requirements gathering, system analysis, design, development, testing, and deployment phases.
* Engineered Python modules to seamlessly extract and load asset data from the MySQL source database, optimizing data retrieval efficiency.
* Utilized the Django framework for application development, ensuring robust and scalable solutions.
* Implemented a file management application capable of downloading, editing, archiving, and publishing files from cloud storage.
* Crafted responsive user interfaces using CSS, HTML, and JavaScript for enhanced website interactivity.
* Designed Snowflake schemas, integral for business intelligence and reporting in OLAP data warehouses, data marts, and relational databases.
* Spearheaded the development of a fully automated continuous integration system using Git, Jenkins, MySQL, and custom Python tools.

**Environment**: Python, R, Artificial Intelligence, Machine Learning, Random Forests and Logistic Regression, Git, Jenkins, MySQL, CSS, HTML, JavaScript.

**Data Torrent Pvt. Ltd, Pune, India**

**Role: Software Developer June 2015 - July 2017**

**Data Torrent enables organizations to accelerate business outcomes through rapid deployment of enterprise hardened, fast big data analytics applications using the latest innovations in operationalization, Data Science and Machine Learning.**

**Responsibilities:**

* Developed and implemented a machine learning library on Apache Apex, a unified batch and stream processing platform for extensive data analysis based on Hadoop
* Integrated MLLib, a machine learning library of Spark, into Apache Apex's engine, to enable running Spark's libraries on Apache Apex's streaming platform.
* Implemented Spark's Lazy Evaluation technique in Java for Apache Apex to enhance the performance of the platform.
* Designed and developed an adapter that enables running Spark's libraries on Apache Apex's engine to facilitate the integration of Spark's libraries with Apache Apex.
* Successfully launched various machine learning algorithms such as SVM, Linear Regression, Logistic Regression, and K-means in Java using Apache Apex's engine.
* Conducted testing and optimization to improve the accuracy and efficiency of the implemented machine learning algorithms on the Apache Apex platform.
* Collaborated with cross-functional teams to ensure adherence to best practices and standards for machine learning in extensive data analysis.
* Contributed to the development of documentation and training materials to facilitate knowledge transfer and adoption of the developed machine-learning library on Apache Apex.

**Environment**: Machine Learning, Apache Apex, Hadoop, MLLib, Spark, SVM, Linear Regression, Logistic Regression.

**TECHNICAL PAPERS & AWARDS:**

* Secured second place in AACoP Hackathon at Abbott for Using Machine Learning to detect Atrial Fibrillation from ECG Signals, San Diego, 2022
* Received Best Idea Award for Complaint Analysis Using Machine Learning, amongst 72 other ideas, at Abbott Rapid Diagnostics, San Diego, 2020.
* Secured First Place in a poster presentation of 85 teams for “Bridging the Communication Gap in Global Fire Safety Using Neural Machine Translation” at Graduate Research Innovation Exchange, WPI, and Worcester 2019.
* Awarded Best Paper Presentation across the Computer Science department with over 20 teams for “Integration of Machine Learning Library in Apache Apex,” Vishwacon, Pune 2017.