

Nitin Solanki

(Senior Full Stack Data Scientist)

Summary

Highly skilled and results-driven Data Scientist with over 9 years of experience in design, development, and deployment of various machine learning, AI, and deep learning solutions. Demonstrated expertise in predictive analytics, data preprocessing, natural language processing, and extracting valuable insights from raw datasets for precise business decision making. Proficient in leveraging advanced techniques and algorithms to solve complex problem statements. Adept at utilizing machine learning, AI, and deep learning frameworks to drive innovation and deliver impactful results. Strong analytical and problem-solving abilities combined with a passion for leveraging cutting-edge technologies in the field of data science.

Technical Skills

Languages & Frameworks: Python, Flask, FastAPI

Databases: SQL, MongoDB, Apache Solr

Libraries: numpy, pandas, scikit-learn, TensorFlow, Keras, gensim, NLTK, spaCy

Machine Learning: Feature Engineering, Exploratory Data Analysis, Feature Selection, Regression, Classification, Dimensionality Reduction, Model Evaluation Techniques, Optimization Techniques, Linear Regression, Naive Bayes, SVM, Logistics Regression, Decision Tree, Random Forest algorithms & Clustering.

Natural Language Processing: Tokenization, Stemming & Lemmatization, Ngrams, POS tagging, Parsing, Language Modeling, Named Entity Recognition, Text-classification, Sentimental Analysis, Question Answering, Word2vec, Stanford NLP, WordNet.

Deep Learning: MLP, CNN, RNN, LSTM, Word2Vec, Fasttext, Encoder-Decoder, Machine Translation, Image Classification, Computer Vision, ELMO, Bert

Generative AI: ChatGPT, open-sourced LLM models such as llama, lamini-lm, dolly etc.

Data Visualization: Matplotlib, Bokeh, Plotly, Tableau

Chatbot Frameworks: RASA, Google dialogflow, Pandorabots(AIML), IBM Watson

Cloud Servers: AWS, Azure, DigitalOcean

Web Crawlers: Selenium, Scrapy

Deployment tools: Docker, Kubernetes
Version Control System: Git

Work Experience

VMWare, Pune

Senior Data Scientist, March 2022 – Present

- Developed a machine learning model to evaluate the risk level of vulnerabilities and prioritize them accordingly, resulting in a more efficient vulnerability management process. This approach has contributed to cost savings of over \$500k annually.
- Created a robust data pipeline for collecting vulnerability feeds from various sources for operating systems and their packages, resulting in streamlined data acquisition processes and containerized the application with docker and orchestrated using Kubernetes.
- Utilized an unsupervised anomaly detection model to classify anomalous event chains for malicious behavior, leading to improved detection accuracy by 30% and reducing potential security incidents, resulting in significant cost savings for the organization.
- Leveraged the pre-trained lamini-lm LLM model to generate question answers on cybersecurity, malwares, and vulnerabilities for the Borathan Event.

AstraZeneca Pharmaceutical, Chennai

Senior Data Scientist, January 2021 – March 2022

- Led a team of NLP specialists in developing and managing data-driven projects that successfully achieved business objectives.
- Fine-tuned the BERT model for Named Entity Recognition (NER) on a small dataset of health regulatory documents, leading to a 40% reduction in manual document review time and significant cost savings in compliance processes.
- Analyzed and classified CSP documents' study design figures using detection2 and image classification techniques, enabling faster access to critical information for decision-making and saving research time.
- Containerized and deployed reproducible deep learning models through docker onto AWS ECR repositories, reducing infrastructure costs by 25% and ensuring scalability and accessibility for the team.

Genpact - Rage Framework Inc., Pune

Data Scientist, October 2019 - January 2021

- Implemented Named Entity Recognition (NER) by classifying various kinds of insurance documents and labeling relevant entities, resulting in significant time savings and cost reductions in data processing.
- Developed an algorithm to detect searchable and non-searchable PDFs, enabling efficient text extraction from non-searchable PDFs and reducing manual efforts.
- Achieved 95% accuracy in classifying insurance documents using a Logistic Regression model trained on word embeddings generated by TF-IDF, leading to faster document processing and saving \$80,000 in manual review costs.
- Deployed TensorFlow models on Amazon SageMaker for inferencing, integrated with AWS Lambda and API Gateway as REST services by optimizing server infrastructure.

Synerzip Softech India Pvt Ltd, Pune

Senior Software Engineer(NLP & Machine Learning), April 2016 - September 2019

- Built a text classification model using the LinearSVC algorithm to categorize phone conversation transcripts between agents and customers, resulting in improved customer service quality.
- Developed and deployed an AWS-based model training pipeline triggered by AWS Lambda functions, ensuring real-time model updates with new data uploads.
- Implemented NLP techniques to analyze and rewrite sentences, including identifying synonyms and converting between active and passive voice.
- Built logistic regression models for predicting diabetes in patients using open-source datasets for proof-of-concept projects.
- Developed and deployed ETL pipelines for data processing, transformation, and loading into AWS Redshift, enabling efficient data analysis and visualization using Python libraries such as Bokeh and Pandas.
- Conducted data preprocessing, cleansing, and feature engineering using NLP techniques, enhancing the quality and usability of the data.
- Utilized various word embedding and dimension reduction techniques to generate training data and improve model performance.
- Developed machine learning models using Python, Spacy, Pandas, Scikit-learn, and TensorFlow/Keras to tackle various data science challenges.

S.S. Userworks Technologies Pvt Ltd/TheLightApp, Pune

NLP Developer, June 2014 - April-2016

- Developed a query correction algorithm using Apache Solr to enhance search accuracy and user experience.
- Implemented a domain detection model to accurately identify the domain of the queried information, improving the relevance of search results.
- Utilized Sparql (DBPedia)/mql (Freebase)/RDF triples to extract subject/predicate/object relations and provide comprehensive answers to queries.

- Conducted web crawling to collect a large amount of data for further analysis.

Accomplishments

- Contributed to Borathan, a global event hosted at VMWare-23, by developing a cutting-edge Cybersecurity GPT for Generative AI and providing valuable support to various teams through LLM models.
- Worked as a Technical Reviewer for the book "Building Chatbots with Python: Using Natural Language Processing and Machine Learning," published by Apress publications. Ensured the accuracy and quality of technical content, contributing to the success of the publication.
- Received recognition and awards for consistently delivering successful AI projects, demonstrating expertise in implementing cutting-edge technologies and achieving desired outcomes.
- Secured the 1st position by showcasing exceptional ideation skills and successfully implementing a search-based application that engages users. Developed a unique Celebrity Matching application based on the Date of Birth (DOB), enhancing user experience and garnering positive feedback from users.

Education

Master of Computer Application(MCA)

University of Pune, 2011-2014

Bachelor of Science(Information Technology)

M.D.S. University, Ajmer, 2008-2011