**Nithin N**

Machine Learning Engineer and Data Scientist

# **Summary**

# Experienced Machine Learning Engineer and Data Scientist with 9 years of experience in the field of machine learning, deep learning, natural language processing, and algorithmic trading. Skilled in designing and implementing ML systems at scale for Websites, Application, Devices, Automation, NLP, chatbots, search, and recommender systems. Proficient in Time-series, Graphs, Geo-tagged, supervised, unsupervised learning and Reinforcement Learning techniques. Strong analytical and problem-solving skills, with proficiency in multiple programming languages and tools.

# **SKILLS**

● **Programming languages:** Python, R, Julia, Scala, Java, C++, C, SAS, Node.js, Next.js, Html, Css, Java script

● **Database technologies:** SQL, NoSQL, MongoDB, Data Bricks, Elastic Search, Redis, Big data, snowflakes, redshift, Teradata, MapReduce

● **Cloud technologies:** Amazon AWS, Docker, MLflow, Airflow, Atlassian tools, AWS S3, Google Cloud Dataflow, Aws Glue, Amazon SageMaker, Emr, Ec2, Lambda, S3, Gcs, Bigquery Databricks, Dataiku, Datarobot, Domino Data Lab, Google Vertex Ai, Ibm, Iguazio, Intel Cnvrg, Microsoft Azure, Paperspace, SAS, Vertex-AI

● **ML Tools:** TensorFlow, Keras, HuggingFace, PyTorch, SpaCy, NLTK, NumPy, Pandas, Scikit-learn, Spark, Kafka,

Matplotlib, PyBullet

● **Software Development:** Algorithms, Flask, RESTful APIs, Docker, Micro Services, Testing

● **MLOPS:** Git, Jenkins, Ansible, pySelenium, Docker, Kubernetes, ArgoCD, Monit, Maven, Powershell

● **Data Visualization:** Seaborn, Plotly, Tableau, Matplotlib, Power BI, QlikView, Communication Skills, Looker, Alteryx, Google Analytics

● **Data Security:** Risk assessment, Vulnerability scanning, Encryption policies and procedures, Encryption configurations, Encryption devices and systems, Remediation planning.

# **EDUCATION**

**Masters in Information Management - Data Science from University of Illinois at Urbana Champaign, USA**

**Bachelors in Electronics and Communications Engineering from GITAM University, India**

**PROFESSIONAL EXPERIENCE**

**R&D Machine learning Engineer and Data Scientist** *Aug 2021 – Present*

*Dovel Technologies, Rockville, MD*

* Developed and maintained **stacked machine learning models and Gradient Boosting Classifier** using **TensorFlow.js**, achieving 99% accuracy in predicting paediatrics patients with Amblyopia. Implemented regularization, hyperparameter tuning, and feature engineering to prevent overfitting and underfitting. Designed and developed a user-friendly UI with an API endpoint for generating reports and integrating the model into applications/websites. Utilized Joblib for efficient data pipelining.
* Designed ML systems and self-running Artificial Intelligence(AI) software to automate Predictive Models.
* Design, Test and Deploy new AI functionalities for various projects of the company.
* Collaborated with doctors and PhDs to collect signal data for disease prediction and developed 9 unique strategies using Python automation. Designed and implemented a streamlined data pipeline using **GitHub, Flask, and Joblib** to extract critical parameters for model training and deployment. Demonstrated a strong ability to work collaboratively with healthcare professionals and leverage technical skills to develop impactful solutions.
* Built and trained RAG models to generate text, translate languages, and answer questions of disease prediction. Successfully applied RAG models to real-world problems, such as generating marketing copy and translating customer support tickets
* Developed and maintained stacked machine learning models with the combination of **decision trees, logistic regression, gradient boost algorithm, Random Forest and neural networks**, I collaborated with a doctor to develop a predictive model for the Parkinson Patients by using Fog and Fall patterns.
* Developed and implemented Tableau dashboards and reports to visualize data and communicate insights to stakeholders.

## R&D Data Scientist *Apr 2019 – July 2021*

*Dovel Technologies, Rock Ville, MD*

* Developed an NLP pipeline using tools like **spaCy and scikit-learn** to perform entity and relation extraction at scale on PubMed and Prostate cancer dataset, later used to automate various processes within MDR and Marketing team.
* Utilized Python with tools such as **scikit-learn, lifelines, and pandas** to design and implement statistical learning, time-to-event, and predictive survival models on real-world evidence (RWE) data sourced from 17M patients’ records.
* Developed a gene-based recommendation model for prostate cancer patients using machine learning techniques and gene analysis, providing treatment insights, with tools including **Python and relevant libraries**.
* Worked with a variety of data sources, including relational databases, flat files, and cloud-based data warehouses.

## Principal Data Scientist *Dec 2016 – Mar 2019*

*American Express, Phoenix, AZ*

* Developed a **Gradient Boosting Classifier ML** model to accurately classify defaulters from a pool of 200,000 customers, identifying 5,000 defaulters with high precision, and designed and implemented a recommendation system to predict future defaulters.
* Developed and deployed machine learning models using Vertex AI. Collaborated with other engineers to build and deploy machine learning models.
* Built Pipelines for data preprocessing using sklearn, feature-engine & featuretools; feature selection/reduction using SelectFromModel, PCA; regression modeling using **GBM, LightGBM, XGBoost**, Random Forest, CatBoost, **Ridge, Lasso and Elastic Net**; hyperparameter tuning using **GridSearchCV, Randomized SearchCV** on **Azure Databricks** in Python for predicting productivity.
* Developed an AI model for customer identity verification using **Natural Language Processing (NLP)** libraries like spaCy and **NLTK, Computer Vision libraries like OpenCV and TensorFlow**, and machine learning frameworks like Scikit-learn and Keras. The model includes text segmentation and extraction techniques that are customized for ID cards, driving licenses, and passports from over 100 countries worldwide. The pipeline time for data scraping has been reduced from one year to less than one month by utilizing optimization techniques such as Gradient Descent and multi-processing libraries like Multiprocessing and Dask.
* Utilized libraries such as **Scikit-learn, TensorFlow, and PyTorch** to conduct advanced analytics, leveraging machine learning algorithms, predictive modeling techniques, and optimization methods to deliver insights into credit risk and develop analytical solutions that help to achieve business objectives.
* Owned and managed dynamic Reject Inference (RI)/Loan Performance data tables (100 million+ rows) from various credit bureaus; Invented and implemented data quality checking, data processing, and updating pipelines using Python libraries such as **Pandas, NumPy, and Dask**, resulting in more efficient and accurate data analysis.
* Developed a deep learning model using face recognition libraries like **OpenFace and dlib**, which replaced the existing API for face verification of customers and resulted in a cost reduction of 15% for the company.
* Optimized the existing model using libraries such as **TensorFlow and PyTorch**, which led to a 25% reduction in the company's core production pipeline time.

## Data Scientist/ Machine Learning Engineer *Dec 2013 – Dec 2015*

***Script Bees Ind Pvt Ltd, India, Hyderabad***

* Designed the product pipeline, and transformed VoIP services through AI to reduce the need for human intervention, providing benefits to businesses with large call volumes.
* Performed in-depth analyses of structured and unstructured data using advanced statistical techniques and strong knowledge of algorithms, resulting in the development of data reports, performance metrics, and strategic objectives to address business partner objectives in **CVS**.
* Performed (**EDA)** exploratory data analysis using pandas-profiling and **K-Means clustering** on NCES data for new business prospects to expand Annuities business.
* Utilized AI/ML and cloud capabilities to build innovative solutions for complex business problems, resulting in improved analytical scenarios and potential future outcomes for 1-Pharma company.
* Developed an algorithm using **OpenCV, YOLO and TensorFlow** for license plate detection which yielded an accuracy of 90%.
* Developed and implemented complex analytical algorithms using MapReduce design patterns. Worked with different varieties of data, including semi-structured and unstructured data, using MapReduce programs and Optimized MapReduce programs for scalability and faster execution.
* Designed a system for 70+ branches of Dominos in the USA, including an algorithm to mask credit card information of customers from call recordings, a sentiment analysis system, an algorithm to detect up-selling, and a pipeline that allows AI-based systems to accept voice or text instructions and perform tasks accordingly.
* Performed data wrangling, manipulation and visualization on 3 million rows of healthcare data for descriptive statistical analysis with **R, Python, Tableau and Excel**.
* As part of my responsibilities, I built an AI/ML enabled Cyber Security product that uses a hybrid neural network (a combination of supervised and unsupervised learning) to detect, handle, and remediate ransomware attacks, while also designing a machine learning pipeline to perform database interactions, data pre-processing, and track model behaviour.