***Tarun Kumar***

***Senior Data Scientist***

***Dallas, TX***

**Email ID:** [**tarunkumarct@gmail.com**](mailto:tarunkumarct@gmail.com)

**Mobile Number: +1 8165962603**

**LinkedIn:** [**https://www.linkedin.com/in/tarun-kumar-1367552a0/**](https://www.linkedin.com/in/tarun-kumar-1367552a0/)

**SUMMARY**

* Around 8 years of Professional IT experience in Analysis, Design, Development, Testing of Enterprise Applications and Database Development.
* Expert in machine learning, statistical analysis, and data visualization, skilled at delivering innovative solutions to address business challenges.
* Adept at leading cross-functional teams, leveraging advanced analytics to optimize processes and enhance organizational efficiency.
* Strong communicator with a knack for translating technical findings into accessible recommendations for senior stakeholders.
* Experienced in developing predictive models and implementing data-driven strategies to propel business growth and competitiveness. Natural Language Processing (NLP): Skilled in NLP techniques, including sentiment analysis, chatbot development, and language translation.
* Computer Vision: Strong background in computer vision applications, encompassing image and video analysis, object detection, and facial recognition.
* Big Data Processing: Well-versed in handling large datasets using tools like Apache Spark and Hadoop, ensuring efficient data preprocessing and feature engineering.
* Model Deployment: Experienced in deploying AI/ML models into production environments, utilizing containerization technologies like Docker and orchestration with Kubernetes.
* AI Ethics and Fairness: Knowledgeable about ethical AI principles and methods for ensuring fairness, transparency, and responsible AI practices.
* AutoML and Hyperparameter Tuning: Familiar with automated machine learning (AutoML) and hyperparameter tuning techniques to optimize model performance.
* MLOps: Proficient in implementing MLOps practices to streamline model development, testing, and deployment.
* Collaborative Team Player: Effective communicator and collaborator in cross-functional teams, translating technical concepts into business insights.
* Continuous Learning: Committed to staying current with the latest AI/ML advancements and best practices.
* Good experience with Amazon Cloud EC2, Simple Storage Service S3, and Amazon SQS.
* Excellent analytical and problem-solving skills and ability to work on my own besides being a valuable and contributing team player.
* Experience with Cloud technologies such as AWS, used EC2, Elastic Block Storage (EBS), Elastic Beanstalk, Lambda, S3
* Experience in developing ETL pipelines in and out of data warehouse using combination of Python and Snowflakes

**EDUCATION**

* **Bachelors** in Electrical Engineering, KL University, 2015
* **Masters** in Computer Science, University of Central Missouri, 2022

**PROFESSIONAL EXPERIENCE**

|  |
| --- |
| *Macy’s Inc., USA* |
| *Senior Data Scientist Oct 2022 - Present* |

Macy's is an American department store chain founded in 1858 by Rowland Hussey Macy. It became a division of the Cincinnati-based Federated Department Stores in 1994, through which it is affiliated with the Bloomingdale's department store chain; the holding company was renamed Macy's. We’re focused on creating compelling experiences to drive traffic, increase engagement and create the potential for new revenue streams in our stores. Developed and supported budget planning and inventory management software for Strategic Planning and Research groups.

**Responsibilities:**

* Proficient Python developer with a strong foundation in data science, using Python as the primary tool for analysis.
* Experienced in data manipulation and analysis with Python libraries like Pandas, NumPy, and SciPy.
* Skilled in data visualization using Matplotlib and Seaborn to communicate insights effectively.
* Proficient in data cleaning, preprocessing, and transformation techniques.
* Proficient in Auto-Sklearn, an automated machine learning (AutoML) library for Python, streamlining model selection and hyperparameter tuning.
* Leveraged Auto-Sklearn to automate the end-to-end machine learning pipeline, reducing development time and improving model performance.
* Demonstrated expertise in Auto-Sklearn for rapid prototyping and model iteration, accelerating time-to-insight.
* Utilized Auto-Sklearn to optimize and fine-tune machine learning models, achieving state-of-the-art performance on various datasets.
* Proficient in developing AI and machine learning solutions using Python.
* Skilled in implementing deep learning models with frameworks like TensorFlow and PyTorch.
* Experienced in natural language processing (NLP) for text analysis and sentiment analysis.
* Strong background in computer vision, including image classification and object detection.
* Utilized scikit-learn for traditional machine learning tasks such as regression and classification.
* Implemented recommendation systems using collaborative filtering and content-based approaches.
* Skilled in data preprocessing, feature engineering, and model evaluation for ML projects.
* Implemented AutoML solutions with Auto-Sklearn to empower non-technical stakeholders with easy access to machine learning.
* Utilized H2O.ai's AutoML functionality to automate model selection, hyperparameter tuning, and feature engineering.
* Implemented distributed computing with H2O.ai's powerful H2O-3 framework, accelerating model training on large datasets.
* Leveraged H2O.ai's XGBoost and GBM implementations for boosting models' predictive accuracy.
* Collaborated with data engineers to integrate H2O.ai into data pipelines, ensuring seamless model deployment.
* Skilled in data visualization to convey insights through storytelling.
* Experienced in crafting data-driven narratives to influence decision-making.
* Expertise in leveraging Python libraries like Matplotlib and Seaborn for data storytelling.
* Proficient in leveraging Amazon SageMaker to build, train, and deploy machine learning models on AWS cloud infrastructure.
* Experienced in building and deploying machine learning models for real-world applications.
* Skilled in data preprocessing, feature engineering, and model evaluation techniques.
* Strong understanding of supervised and unsupervised learning algorithms.
* Experienced in handling missing values, outliers, and duplicate data for improved data quality.
* Skilled in data normalization, standardization, and transformation for analysis-ready datasets.
* Implemented metadata tracking systems to capture and manage data lineage information.
* Ensured data lineage transparency and accountability, enabling confident decision-making.
* Leveraged Python to interact with Amazon Redshift, automating data pipelines.
* Enhanced query performance by optimizing and tuning Amazon Redshift clusters.
* Implemented ETL processes using Python and Amazon Redshift for data transformation.
* Created and managed Redshift data warehouses to support analytical workloads.
* Experienced in optimizing SQL queries for BigQuery to improve query performance.
* Skilled in designing and managing data pipelines that utilize BigQuery as a backend data store.
* Developed Python-based Hadoop applications for distributed data processing, improving data scalability and performance.
* Implemented data ingestion pipelines using tools like Apache Flume and Sqoop to seamlessly transfer data into Hadoop clusters.
* Skilled in using Matplotlib to communicate complex data insights effectively.
* Experienced in customizing Matplotlib plots for publication-quality graphics.
* Adept at creating a wide range of plots, including line plots, bar charts, scatter plots, and heatmaps using Matplotlib.
* Skilled in using Cloud Functions and Cloud Run for serverless Python application development.
* Review code and system interfaces and extracts to handle the migration of data between systems/databases.
* Experience working with Amazon Web Services (AWS) for improved eﬀiciency of storage, Security, and fast access.
* Facilitate creation of the tasks for the sprint with dependencies is identified. Work with key Subject Matter Expertise and partner to troubleshoot and prioritization actions
* Experience in designing the Blue Prism process solutions inline the Blue Prism template and best practices.
* Experience in creating process flows using Process Studio, Object Studio and Control Room in Blue Prism
* Expertise in using Application Modeller, Spy Modes and Attributes
* Created end to end automation solution for end users as per requirement which involves feasibility study, development, testing and deployment
* Involved in developing APIs for Connection to Elastic search using Jest.
* Used TDD/BDD with unit testing using Junit, Mockito and Karma
* Implemented the Agile Scrum methodology and Test-Driven Development (TDD) for developing the application
* Created a Spark cluster on AWS EC2 and integrated with IPython to provide team with machine learning environment.
* Implement new cloud management platforms from AWS, OpenStack, Google Cloud Platform.
* Conducted version control, software maintenance and backup using Maven and Git.
* Configured TFS SharePoint Services and Reporting Services, created Project portals for all existing Team Projects.
* During the life cycle of the project my responsibility also includes MySQL database administration and maintenance over Linux server.
* Configured the Slack in Jenkins and published the build status notifications.
* Docker Jenkins with Master and Slave architecture in OpenShift platform and automated the build jobs.
* Used Kubernetes to deploy scale, load balance, and manage Docker containers with multiple namespace versions.

**Environment- Data Science, Auto-Sklearn, H2O.ai,Data Storytelling, AWS, SageMaker, Google AI Platform, Azure, Machine Learning, machine learning models, Data cleaning, data lineage, data cataloging , Python, BERT, GPT (Generative Pre-trained Transformer), Word Embeddings, Named Entity Recognition (NER),OpenCV, YOLO, OpenAI Gym , Deep Q-Networks (DQN), Policy Gradients, PyTorch, scikit-learn, NumPy, Es6, typescript, JSON web token, Java, Postgres, MySQL, SQL, PL/SQL, CSV, PostgreSQL, Cassandra, Rest, PyCharm, GitLab, Git.**

|  |
| --- |
| *Saber Health Group, USA* |
| *Senior Data Scientist Jan 2022 - Oct 2022* |

Saber Health Group delivers exceptional healthcare services including rehabilitation, skilled nursing and long-term care. Saber Healthcare Group, LLC owns and operates skilled nursing and rehabilitation facilities for residents and their families.

**Responsibilities:**

* Strong knowledge of machine learning algorithms and libraries such as Scikit-Learn.
* Hands-on experience with deep learning frameworks like TensorFlow and Keras.
* Applied statistical analysis to derive actionable insights from data.
* Developed predictive models for various business applications using Python.
* Employed Auto-Sklearn's ensemble techniques to improve model robustness and generalization.
* Integrated Auto-Sklearn into production workflows for automated model retraining and deployment.
* Collaborated with cross-functional teams to deploy Auto-Sklearn-based models in real-world applications.
* Conducted anomaly detection tasks using H2O.ai's Isolation Forest and AutoEncoder algorithms.
* Employed H2O.ai's NLP capabilities for text data analysis and sentiment analysis tasks.
* Contributed to the H2O.ai community by sharing custom model recipes and participating in forums.
* Strong ability to communicate complex findings in a clear and concise manner through data storytelling.
* Proven track record of translating technical analysis into actionable business recommendations.
* Skilled in data preprocessing and feature engineering using SageMaker Processing jobs, ensuring high-quality input for model training.
* Proficient in implementing computer vision and natural language processing (NLP) tasks.
* Developed and fine-tuned deep learning models for image classification and object detection.
* Utilized Python libraries like Pandas and NumPy for efficient data cleaning and manipulation.
* Utilized data lineage insights to improve data quality through identification of data flow issues.
* Designed and maintained data models in Amazon Redshift for reporting and analysis.
* Improved data security in Redshift by setting up IAM roles and encryption.
* Accomplished in using Python libraries like pandas and matplotlib to analyze and visualize BigQuery data.
* Utilized Hadoop's YARN resource manager to optimize cluster resource allocation and enhance job execution.
* Knowledgeable in interactive plotting with Matplotlib's interactive mode.
* Skilled in working with large datasets and distributed computing frameworks like Apache Spark.
* Experienced in A/B testing and experimentation to measure the impact of ML models on business metrics.
* Expertise in building and deploying H2O.ai models for real-time predictions and batch processing in Python.
* Implemented interpretable machine learning models with H2O.ai's Driverless AI, enhancing model transparency and trust.
* Developed backend components using Python Django framework.
* Good working knowledge in Python Object Oriented Programming Concepts (OOP’s).
* Good working experience with Python libraries NumPy, Pandas, matplotlib.
* Utilized Python libraries NumPy and matplotlib for data analysis and data manipulation.
* Built various graphs for business decision making using Python matplotlib library.
* Experience using Content Management Systems: WordPress and Django CMS plugins.
* Supported/maintained client website within the DJANGO CMS system for various website needs.
* Worked on Django Permissions while developing Python API
* Knowledgeable in edge AI device selection and integration with Python-based AI solutions.
* Competent in leveraging TensorFlow Lite and PyTorch Mobile for edge AI deployments. Proficient in deploying TensorFlow models to production environments, including cloud platforms like Google Cloud AI Platform.
* Utilized pre-trained models from TensorFlow Hub and fine-tuned them for specific tasks, saving significant training time and resources.
* Skilled in leveraging PyTorch's flexibility and efficiency for building custom neural networks.
* Proficient in leveraging MXNet's GPU acceleration for accelerated model training and inference.
* Implemented various neural network architectures, including feedforward neural networks (FNN), CNNs, RNNs, and transformer models.
* Skilled in fine-tuning pre-trained BERT models for various NLP applications.
* Utilized GPT models for automatic text summarization, making lengthy documents more digestible.
* Developed NER pipelines for automating the extraction of entities such as names, dates, locations, and more from text.
* Adept at using OpenCV to develop custom image processing pipelines for specific project requirements.
* Strong expertise in fine-tuning YOLO models for custom datasets and specific tasks.
* Adept at integrating OpenAI Gym with popular RL libraries like TensorFlow and PyTorch.
* Skilled in handling and cleaning messy data with Pandas, ensuring data quality and accuracy.
* Enhanced data visualization efficiency by customizing Matplotlib's styles and themes.
* Familiar with Google Cloud's Big Data services, including BigQuery and Dataflow, for handling large datasets with Python.
* Created database tables in MySQL for ORM mapping
* Contribution in development of workflow engine as a AOP Dockers Spring Security Microservices framework using drools.
* Experienced in JHipster stack for building monolithic as well as microservices based architecture.
* Involved in developing a Restful service using Python Flask framework.
* Used Celery as task queue and RabbitMQ as messaging broker to execute asynchronous tasks.
* Create reusable processes and/or objects for Blue Prism
* Prepared RPA program by establishing roles and responsibilities, coaching and supporting RPA training, and agile mentoring
* Created AWS RDS and mapping the EC2 JBOSS configurations to point to AWS RDS instance.
* Worked on CI/CD pipeline and created sandbox, UAT and Production Environments in Google Cloud Platform.
* Prepared Java/J2EE development structure for XCode, Maven.
* Configured SharePoint foundation services with TFS
* Lightweight virtualization with Linux Containers and Docker
* Configured Alert manager to send alerts to Ops Genie and HipChat.
* Worked on Microservices for Continuous Delivery environment using Docker and Jenkins.
* Used Jenkins pipelines to drive all micro services builds out to the Docker registry and then deployed to Kubernetes.

**Environment** - **Data Science, Auto-Sklearn, H2O.ai,Data Storytelling, AWS, SageMaker, Google AI Platform, Azure, Machine Learning, machine learning models, Data cleaning, data lineage, data cataloging , Python, BERT, GPT (Generative Pre-trained Transformer), Word Embeddings, Named Entity Recognition (NER),OpenCV, YOLO, OpenAI Gym , Deep Q-Networks (DQN), Policy Gradients, PyTorch, scikit-learn, NumPy, Es6, typescript, JSON web token, Java, Postgres, MySQL, SQL, PL/SQL, CSV, PostgreSQL, Cassandra, Rest, PyCharm, GitLab, Git.**

|  |
| --- |
| *C2N IT Services, India* |
| *Data Science and Engineer Jan 2019 – Nov 2020* |

**Responsibilities:**

* Worked as dev-ops create deployment strategy and scripts (BASH/Python)
* Developed scripts for build, deployment, maintain and related task using Jenkins, Python.
* Developed web applications by following Model View Control (MVC) Architecture using server-side applications Django, Flask and Pyramid.
* Added features to and maintained existing Django and Pyramid webapps.
* Performed CRUD operations on MySQL database using Django's ORM.
* Created automated scripts for incremental data loading in SQL Server using Rest API and Pandas
* Developed API services in Python/Tornado while leveraging AMQP and RabbitMQ for distributed architectures.
* Developed and tested many features in an AGILE environment using HTML5, JavaScript and Bootstrap
* Windows Server machines running PostgreSQL DBMS, presented with HTML5 and CSS3
* Applied JQuery scripts for basic animation and end user screen customization purposes.
* Experience working with IBM Cast Iron management APIs to access the Cast iron WMC features.
* Experience configuring the Cast Iron WMC.
* Worked as Cast Iron developer and supported fixing bugs across Cast Iron product, developing new features and helping with analysis and fixing the PMRs for the runtime component and SFDC for all forms factors of IBM WebSphere Cast Iron.
* Worked as dev-ops create deployment strategy and scripts (BASH/Python)
* Developed scripts for build, deployment, maintain and related task using Jenkins, Python.
* Developed web applications by following Model View Control (MVC) Architecture using server-side applications Django, Flask and Pyramid.
* Added features to and maintained existing Django and Pyramid webapps.
* Created automated scripts for incremental data loading in SQL Server using Rest API and Pandas
* Developed API services in Python/Tornado while leveraging AMQP and RabbitMQ for distributed architectures.
* Developed and tested many features in an AGILE environment using HTML5, JavaScript and Bootstrap
* Windows Server machines running PostgreSQL DBMS, presented with HTML5 and CSS3
* Applied JQuery scripts for basic animation and end user screen customization purposes.

**Environment- Data Science, Auto-Sklearn, H2O.ai, Data Storytelling, AWS, SageMaker, Google AI Platform, Azure, Machine Learning, machine learning models, Data cleaning, data lineage, data cataloging, Python, BERT, GPT (Generative Pre-trained Transformer), Word Embeddings, WPF, Net, DevExpress, DAPI, Autosys, OMS, PROD, CyberArk, UAT, Operating System, Linux, Windows, Chrome DT**

|  |
| --- |
| *Larsen & Toubro Tech Services, India* |
| *Python developer Feb 2017–Jan 2019* |

**Responsibilities:**

* Familiar with natural language processing (NLP) for text data analysis.
* Experience in feature engineering to improve model performance.
* Built and deployed machine learning models in production environments.
* Strong understanding of databases, including SQL and NoSQL for data storage and retrieval.
* Enhanced model interpretability through Auto-Sklearn's feature importance analysis, making complex models more understandable.
* Developed custom machine learning pipelines using Auto-Sklearn's flexibility and extensibility for specific project requirements.
* Applied H2O.ai's explainability tools to interpret complex model decisions and enhance model transparency.
* Enhanced model performance and scalability using H2O.ai's GPU support for deep learning models.
* Adept at creating interactive data dashboards for intuitive storytelling using tools like Plotly and Dash.
* Experienced in developing, optimizing, and fine-tuning machine learning models using SageMaker's Jupyter Notebook integration.
* Demonstrated expertise in handling categorical data through encoding and feature engineering.
* Conducted impact analysis using data lineage to assess the consequences of data changes.
* Utilized Redshift Spectrum to query data in Amazon S3 directly from Redshift.
* Familiar with Google Cloud's Data Studio for creating data visualizations based on BigQuery data.
* Designed and maintained Hadoop cluster environments, ensuring high availability and reliability.
* Proficient in integrating Matplotlib with other Python libraries like NumPy and Pandas for seamless data visualization.
* Involved in designing a data management system using PostgreSQL, DynamoDB.
* Designed and developed KML files for Plotting Information on Google Earth and Google maps.
* Designed and developed debugging tools for testing, Firmware updates, etc.
* Formulated SQL queries, Aggregate Functions, and database schema to automate information retrieval.
* Developed a fully automated continuous integration system using Git, Gerrit, Jenkins, MySQL and custom tools developed in Python and Bash.
* Converted Visual Basic Application to Python, MSQL.
* Used python scripts to update content in the database and manipulate files.
* Created Reports using JASPER.
* Used Thales asynchronous theorem for applying encryption and decryption of ISO standard message in Python programming.
* Developed the tools using Python Django and used MongoDB for databases. Parsers written
* in Python for extracting useful data from the design data base. Used Parse kit (Enigma.io)
* framework for writing Parsers for ETL extraction.
* Experience in working with Drag and Drop Calculation, and Geographic Search by using Tableau.
* Spark Streaming collects data from Kafka in near-real-time and performs necessary transformations and aggregation to build the common learner data model and stores the data in NoSQL store (MongoDB).
* Experienced in developing Web Services with Python programming language. Good experience with Unix, Linux scripting/unit testing/Load testing, Performance stress/endurance testing and Integration Testing; tools and frameworks- Junit, PyUnit, HP VUGEN LoadRunner, Selenium Web driver, Controller, Dynatrace, Performance Center, QC, JIRA.
* Build back-end application with Python / Django, work with Docker, RabbitMQ, Celery, Jenkins.
* Used data types like dictionaries, tuples and object-oriented concepts-based inheritance features for making complex algorithms of networks.
* Experience in debugging and troubleshooting programming related issues.

**Environment:** **Python, CSS, JQuery, JavaScript, Apache, Ajax, simplemodal.js, Git, PowerShell, Selenium, Linux, Bootstrap, GitHub, Django, VBscripts, Jasper, Gerrit, MySQL, Zookeeper, Unix, HTML.**

|  |
| --- |
| *PayPal Data Services, India* |
| *Python developer May 2015- Feb 2017* |

A dispute management tool developed by PayPal to log, manage, and adjudicate conflicts that arise between end users and the merchants.

**Responsibilities:**

* Extensive experience in designing, developing and enhancing applications using PERL, Shell and Python scripting.
* Designed and built component-oriented libraries and modules for internal tools using Python/Unix.
* Developed Python webapps using Django and Pyramid to communicate with internal systems.
* Used Python to write data into JSON web token files for testing Django Websites & Created scripts for data modelling and data import and export.
* Experience in working with Python ORM Libraries including Django ORM.
* Used Pandas library for statistical Analysis.
* Strong Core Java background with experience in Collections, Multi-Threading, Java 8, Interfaces, Serialization, Synchronization, Exception Handling, OOP's techniques.
* Proficient in using the IDE like Eclipse for developing the Java applications.
* Experience in writing the SQL scripts like DML and DDL by using the CRUD operations and using the various joins in MySQL and Oracle SQL.
* Experience in using version control tool like GIT.
* Automate both the GUI and runtime testing using tools RFT and Junit
* Develop java application using RAD and deploy on standalone and multimode server on multiple Operating systems.

**Environment**: **Python, MySQL, HTML, JavaScript, PHP, PostgreSQL, Unit test, JDBC, JSP, Servlets, Frameworks, Spring MVC, Spring Boot, Django.**