VIKASH KUMAR

Delhi NCR *Phone: 9205083231*
linkedin.com/in/vikas-kr-76642a175 *E-mail: vickynicky.99@gmail.com*
https://github.com/Vicks-2019

Process-oriented Data Scientist with hands-on experience in data interpretation, visualization, statistical modeling, and advanced analytics. Skilled in building machine learning and Gen AI solutions to solve real-world business problems and support data-driven decision making. Looking for a Data Scientist role to leverage analytical and technical competencies in driving impactful outcomes.

Programming: Python, R, R-Shiny, MySQL, HTML, CSS, JavaScript, React.js, Flask**,** Fast API
ML / DL Packages: Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn, NLTK, Spacy, OpenCV, TensorFlow, Karas, PyTorch
Tools / Platforms: Tableau, Power BI, Adv Excel, GitHub, AWS ,Azure, MLOps , LangChain, LangGraph , llamaindex ,OpenAI
Techniques: Machin Learning, Deep Learning, Regression, Classification, Clustering, Segmentation, Time Series Forecasting, Probability & Statistics, Neural Network, NLP, LLM, Generative AI, RAG, Agentic AI, vector dB , Agentic security

PROFESSIONAL EXPERIENCE

Estater Pvt Ltd (GCC) Oct 2022 – Present
Data Scientist

- Working on SAS-based analytics product and collaborating cross-functionally to identify insights and resolve issues.
- Developed ML models for real estate valuation to enable investment decisions and maximize accuracy in production.
- Analyzing Data to figure out features related to property and building machine learning models to predict value of property, ensuring best result on production that helps clients in decision making for investment.
- Enforced alignment of project strategy with business objectives and made modifications to promote efficient project completion & Forecasting.
- Contributed to new product development using NLP & Generative AI within the real-estate domain.

Precognitas health Pvt Ltd (Foresight Health Solutions, A US based start-up)
Data Scientist & Python Developer Dec 2021 – Sept 2022

- Designed analytics products using Python, ML & NLP, processed healthcare datasets via Python, SQL, AWS & BI tools.
- Built Risk & Intervention models to optimize healthcare costs using ML/DL techniques and deployed solutions.
- Conducted statistical modeling based on client requirements and ensured continuous improvements.
- statistically interpret key points from gathered data. Investigated and corrected as per client requirements.

ValuemaxIndia Jul 2020 – Dec 2021

Data Science Developer

- Understanding data, performing predictive analysis using python & Machine learning
- Developed Machine Learning & Deep Learning applications using Flask/Django & deployed on Heroku.
- Built multiple modern web apps using ReactJS (hooks, axios, redux) with Python backend connectivity.
- Performed end-to-end data processing including cleaning, preparation & visualization with BI dashboards.

Asset plus consulting Jun 2019 – Jun 2020
Data Analyst

- Conducted data preprocessing, visualization, and predictive analysis using R and Python.
- Built consumer payment behavior predictive model for DISCOM to improve on-time revenue collection.
- Implemented NLP-based solution for E-tagging to flag fraudulent connections.
- Created advanced Tableau dashboards using joins, union, blending & multiple chart types.

Kotak Mahindra Bank Ltd Sep 2018 – Mar 2019
Assistant Manager

- To prepare MIS report weekly & monthly as per set norms using MS Excel, MS Doc.
- To handle the business case and ensure proper solution to corporates by analyzing the data.
- To make synergy with the other verticals of the organization in order to prepare the report.
- To prepare report to help with auditing and hence enhancing the organization performance.

STEAG Energy Services (India) Pvt Ltd Oct 2012 – Jun 2016
Planning & Operation Engineer

- Analyze key metrics, identify key trends and prepare reports and presentations for management
- Developed visualization reports for daily, weekly, monthly and yearly conversion for timely comparison.
- Complete requirement analysis, trouble shooting and finding RCA of issues for the test & Production environment.
- Maintaining 5s in the area & Evaluating Data sources and strong understanding of Data warehouses

---

EDUCATION

- PG Diploma Certification in Data Science & Machine learning (Data Trained) in 2019-2020
- MBA (Finance & Marketing) from CMS (Centre for Management Studies, JMI) in session 2016-2018 with 8.0 CGPA
  B.E. in Electrical and Electronics engineering from VTU University Bangalore with 70%.

PROJECTS

Brochure Digitization-RAG-Based PDF Data Extraction (Estater)

- Built a secure **RAG-based pipeline** to extract structured data from unstructured PDFs using LLM-driven semantic retrieval.
- Implemented data masking, PDF text extraction, chunking, and embedding generation with Sentence Transformers, storing vectors in FAISS.
- Designed schema-driven prompts to generate consistent JSON outputs, directly convertible into analytical data frames.
- Enabled context-aware extraction via LLM reasoning, improving automation and accuracy of document processing
- **Tech:** Python, Lang Chain, FAISS, Sentence Transformers, Hugging Face LLMs, Regex, JSON

Rental Price Prediction – Apartments & Villas (Client – NHC Saudi Arabia)

- Developed **ANN-based rental price prediction models** for apartments and villas to support real-estate valuation decisions.
- Performed end-to-end data preprocessing including missing value handling, outlier removal, encoding, and feature scaling.
- Engineered domain-specific features such as price per sqft and property age buckets to improve prediction accuracy.
- Trained and optimized models using ReLU, dropout, Adam optimizer, and early stopping**,** benchmarking performance against traditional ML models using RMSE, MAE, and R^2.

Multilingual Speech-to-Text (Estater)

- Developed a **speech-to-text data extraction** pipeline for English and Arabic audio inputs.
- Cleaned, normalized, and processed transcribed text to extract structured insights.

- Implemented fuzzy matching and rule-based logic to identify key entities from speech data.
- Integrated CI/CD workflows to automate deployment and updates.
- Tech : Python, Fast API, NLTK, Regex, Fuzzy Wuzzy, Translation APIs

Brand Data Generation Using Generative AI (Client – NHC Saudi)

- Built a Generative AI solution to automatically generate brand-related structured data using prompt driven LLMs.
- Designed prompts to ensure consistency, formatting accuracy, and business relevance.
- Validated outputs using rule-based checks and semantic similarity measures to maintain data quality.
  Tech : Python, Lang Chain, Prompt Engineering, LLMs

Land Parcel Rate Prediction (Estater)

- Developed **Automated Valuation Models (AVM)** to predict land parcel rates using statistical and machine learning techniques.
- Conducted EDA, feature analysis, and model comparisons to identify the most accurate valuation approach.
- Automated the valuation pipeline and delivered model insights and performance reports to support stakeholder decision-making. R, RStudio, Statistical Modeling, Regression.

Patient Risk & Intervention Modeling – (Client-WCCHC (USA))

- Developed **patient risk prediction models** using EHR, claims, and demographic data to identify highcost patients.
- Built an intervention recommendation model using ML/DL to suggest cost-effective care actions and early interventions.
- Enabled better resource allocation, reduced healthcare costs, and improved patient outcomes through predictive and prescriptive analytics. **Tech** Python, Machine Learning, Deep Learning, ANN.

Car Price Prediction System (*Client: Sriram Finance*)

Developed machine learning models to estimate car prices using features such as brand, model, year, mileage, fuel type, and condition.

- Trained and evaluated models using **Linear Regression and Decision Trees** to identify optimal pricing performance.
- Deployed models via Flask/FastAPI APIs for real-time predictions, with continuous monitoring and retraining to ensure accuracy in dynamic market conditions.

E-Tagging for Fake Consumer Detection (Client – BSES Delhi)

- Built an **NLP-based fraud detection system** to identify fake utility consumers using unstructured text data (addresses, forms, complaints).
- Cleaned and processed text, extracted entities, and identified suspicious similarities and duplicate patterns.
- Classified consumers as genuine or fake, helping reduce fraud, automate verification, and improve operational efficiency.

  Tech **:** Python, scikit-learn, spaCy, NLP, Machine Learning.