PRABAL KUMAR DEKA

Phone: +918638517995 | Email: prabal456.pkd@gmail.com | Guwahati, Assam| LinkedIn | GitHub

SUMMARY

Entry-level Data Scientist with proven experience in Machine Learning, Deep Learning, and Natural Language Processing (NLP). Expertise in Python, SQL, and statistical analysis with strong problem-solving abilities and experience in delivering data-driven insights for real-world applications.

PROFESSIONAL EXPERIENCE

Data & Research Analyst – Turing **Sept 2024 – Aug 2025**

- **Data Annotation & Quality Assurance:** Collaborated with cross-functional stakeholders to define requirements for large-scale data annotation projects, implementing quality control measures that achieved 90% annotation accuracy across 50K+ data points.
- **Large Language Model (LLM) Optimization:** Fine-tuned and evaluated pre-trained language models , achieving 95% accuracy improvement in mathematical reasoning tasks and reducing inference time by 40%.

Computer Science Faculty – Potential & Concept Educations LLP **Mar 2023 – Sept 2024**

- **Curriculum Development:** Designed and delivered comprehensive programming curricula for 100+ students in Class 11 & 12, covering Python fundamentals and database management (SQL queries, joins, views, CTEs).
- **Technical Training:** Conducted hands-on workshops on data manipulation and analysis using Python libraries, resulting in 90% student pass rate in board examinations.

PROJECTS

Cars 24x7: Car Damage Detection (Link)

- Built and compared **CNN, EfficientNetB0, and ResNet50** models to classify car damages into 6 categories.
- Optimized **ResNet50** using **Optuna** for **hyperparameter tuning**, achieving **79% accuracy** on test.
- Developed a **Streamlit web app** for real-time image-based damage detection.

LinkedIn Post Generator – Generative AI Project (Link)

- Developed an end-to-end Generative AI application using **LangChain** framework and **Llama-3.3** model for automated content generation.
- Implemented prompt engineering techniques and fine-tuning strategies to improve content relevance by 85%
- Built interactive web interface using **Streamlit** with real-time content generation capabilities.

Healthcare Premium Prediction – Machine Learning Project (Link)

- Built predictive models to estimate healthcare insurance premiums using ensemble methods (**Random Forest, XGBoost, Gradient Boosting**).
- Performed feature engineering on dataset with **10K+** records, achieving **92%** prediction accuracy.
- Implemented cross-validation and hyperparameter tuning using **GridSearchCV** to optimize model performance.

Credit Risk Modelling – Prudent Partners (Link)

- Developed machine learning models (**Logistic Regression, Random Forest, XGBoost**) to assess credit default risk for loan applications
- Utilized feature engineering techniques and handled class imbalance using **SMOTE** and ensemble.
- Implemented hyperparameter optimization using **Optuna**, improving model **AUC** score to **0.89**.

TECHNICAL SKILLS

- **Programming Languages:** Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), SQL (Joins, CTEs, Window Functions, Stored Procedures)
- **Machine Learning:** Supervised/Unsupervised Learning, Regression, Classification, Clustering, Feature Engineering
- **Statistical Analysis:** Hypothesis Testing, A/B Testing, Descriptive & Inferential Statistics
- **Deep Learning:** Neural Networks, Convolutional Neural Networks (CNN), Transfer Learning, PyTorch.
- **Natural Language Processing (NLP):** Text Preprocessing, Sentiment Analysis, LangChain
- **Data Visualization:** Power BI (DAX, Power Query)
- **Databases:** MySQL
- **Generative AI:** Large Language Models (LLMs), Prompt Engineering, Fine-tuning, RAG
- **MLOps:** Model Deployment, FastAPI, Git
- **Development:** Jupyter Notebook, PyCharm, VS Code, Google Colab
- **Business Intelligence:** Advanced Excel (Pivot Tables, Power Pivot, VLOOKUP)

EDUCATION

- **Central Institute of Technology, Kokrajhar, Assam** June 2018 – June 2022
  B. Tech in Computer Science & Engineering.
- **GATE 2023 (CSE) Qualified**

CERTIFICATIONS

- Machine Learning for Data Science – Codebasics.io
- Deep Learning: Beginner to Advanced - Codebasics.io
- Python for Data Professionals – Codebasics.io
- SQL for Data Analysis – Codebasics.io

ACHIEVEMENTS

- **Model Accuracy Improvement:** Achieved 95% accuracy improvement in LLM fine-tuning for mathematical reasoning tasks at Turing.
- **Research Impact:** Contributed to 3 data annotation projects with 95%+ accuracy standards at Turing.
- **Teaching Excellence Award:** Maintained 90% student success rate in computer science coursework at Potential & Concept Educations.

LANGUAGES

- English: Professional Working Proficiency
- Hindi: Native/Bilingual Proficiency