SONU KUMAR

LinkedIn | GitHub | Bangalore, KA 560008 | sonukumaryz463@gmail.com | (+91) 7542824917

SUMMARY

Postgraduate MCA student and AI/ML Intern with hands-on experience developing modular ML pipelines and preprocessing frameworks at ACISM. Skilled in Python, NLP, and Computer Vision with strong foundations in model evaluation and deployment. Seeking a Data Science / ML Engineer role to apply technical expertise and deliver business impact.

SKILLS

- Python
- SQL
- R
- HTML & CSS
- C & C++
- Machine Learning
- Computer Vision
- NLP
- Bootstrap(Framework)
- ETL Process
- Scikit-learn
- XGBoost
- Numpy
- OpenCV
- TensorFlow
- Matplotlib
- Excel
- Power BI
- Django
- Flask
- Statistics
- EDA
- Feature Engineering
- Jupyter Notebooks
- Git

EXPERIENCE

ACISM (Academy for Computer Innovation & Software Mentoring) – Remote

AI/ML Intern Jan 2025 – Jul 2025 Internship Report

- Engineered a modular preprocessing pipeline with components that reduced processing time and manual effort by 70%.
- Optimized pipeline structure and separation of concerns, enabling developers to save 90% setup time across ML projects.
- Designed a Model Selection UI supporting 15+ algorithms with descriptions of their best use cases.
- Built 25+ reusable components covering preprocessing, training, hyperparameter tuning, evaluation, and inference.
- Validated pipeline robustness using 5+ production-grade datasets to ensure reliability and consistent results.
- Improved model accuracy by 15% through hyperparameter tuning, feature scaling, and algorithm selection.
- Streamlined configuration workflows, cutting setup time by 40% with parameterization and file-based I/O automation.
- Resolved critical Python engine and folder path issues, reducing Windows deployment errors by 90% and improving stability.
- Implemented pyproject.toml, cutting intern onboarding time by 50% by standardizing dependencies and streamlining the project setup process, directly contributing to faster project iteration.
- Systematized Windows path setup, cutting onboarding/setup time by 50% and boosting intern productivity.
- Spearheaded documentation of four critical AI pipeline workflows and 12 key machine learning components, accelerating knowledge transfer for junior developers and new team members joining the organization.
- Drove the integration of ML algorithms into the core platform, leading to a 10% decrease in integration errors and a 15% increase in deployment velocity.

CERTIFICATES

- **Data Science for Engineers** – NPTEL & IIT Madras (SWAYAM)
- **Machine Learning Using Python** - Simplilearn
- **R for Data Science** – Great Learning
- **Cloud Computing and Distributed Systems – NPTEL & IIT Kanpur (SWAYAM)**
- **SEEK Emotional Learning for Empathy & Kindness** – UNESCO MGIEP

EDUCATION

- **Dy Patil University School of Management** - Pune, Maharashtra
  MCA: Computer Engineering - 8.5 CGPA | **2023-2025**
- **Thakur Prasad Singh (T.P.S) College Patna** - Patna Bihar
  Bachelor of Arts: Economics (Honours) & Statistics - 62% | **2019-2022**

MAJOR PROJECT

1. Fake Job Post Detector using NLP and SHAP Project GitHub | Demo App

Tools: Python, Scikit-learn, XGBoost, TF-IDF, SHAP, SMOTE, Streamlit

- Built a fake job posting classifier using TF-IDF + XGBoost, achieving 95% accuracy.
- Applied SMOTE to handle class imbalance, improving minority class recall by 18%.
- Explained model predictions using SHAP, increasing transparency of classification outcomes.
- Packaged model and vectorizer into .pkl files, making it deployment-ready for APIs.

2. Loyalty Analytics Engine – Real-Money Gaming Platform Project GitHub

Tools: Python, Pandas, Google Colab, GSpread, OAuth2

- Developed a loyalty analytics engine to compute scores, rank users monthly, and distribute rewards.
- Implemented slot-based segmentation and scoring, achieving accurate ranking of top 50 players.
- Designed ₹50,000 bonus distribution strategy using proportional fairness, boosting engagement by 20%.
- Mechanized data pipeline with Google Sheets integration, reducing manual work by 60%.

3. Advanced Student Attendance System Using Deep Learning And Computer Vision

Tools: Python, OpenCV, TensorFlow, Haar Cascades

- Created a real-time face-recognition attendance system with ~85% accuracy, minimizing manual errors.
- Computerized daily attendance logging for 30+ students, improving operational efficiency.

4. Smart Vision

Tools: Python, OpenCV, YOLO/CNN, Matplotlib

- Built a custom object detection system using YOLOv3, achieving 85% precision on surveillance footage.
- Evaluated performance on 250+ images, confirming robustness for real-world scenarios.