# Saugat Thapa Chhetri

Nagarjun-09, Kathmandu, Nepal | saugat556513@gmail.com | Ph: +977 9864985654 | linkedin:Saugat Thapa Chhetri | Portfolio:404saugat404.com

## ACADEMIC QUALIFICATION

Tribhuvan University, National College of Engineering , Lalitpur
**Bachelor in Computer Engineering** *2019 - July, 2024*

## SKILLS,ACHIEVEMENTS AND TRAININGS

- **Data Science & ML Libraries:** Scikit-learn, NumPy, Pandas, Matplotlib, SciPy, XGBoost, PyTorch, PyTorch Lighting
- **AI agents:** HuggingFace, Langchain, Langgraph, Transformers, Ollama, n8n, Groq, LamaIndex
- **Database and Big Data:** PostgreSQL,kafka,airflow(learning), Milvus, Quadrant, ChromaDB
- **Web Development:** FastAPI(Backend),React(Frontend), Git(Version Control), GitHub,Keycloak(security).Gitlab
- **Machine Learning Areas, Visualisation & Reporting: :** Supervised & Unsupervised Learning, Time Series Forecasting, Anomaly Detection, Model Evaluation & Tuning,Seaborn, Matplotlib, Streamlit, Tableau (basic)
- **Fusemachine AI Fellowship (Apr - Oct 2024):** Competitive 6-month AI/ML fellowship; built and deployed churn, forecasting, and classification models and delivered a capstone project.

## EXPERIENCE

### MetaCloud Solutions

**AI,ML Engineer** *May 12, 2026- Present*

- Built a **real-time voice AI bot** for telephony, integrated with an **Asterisk** server, with agentic tool-calling and live call transfer to human agents, delivering responses in **under 1.8 seconds** across the [ASR → LLM → TTS] pipeline.
- Built an agentic customer-support chatbot for **Worldlink** using LangGraph and FastAPI, with tool-calling agents that handle [e.g., plan queries, billing, complaints] and escalate to human agents when needed.
- Developed a **customer AI insights dashboard** with a FastAPI and PostgreSQL backend, using a **nightly cron job** that extracts and processes customer data to generate AI-driven insights on [churn, usage trends, customer behavior], visualized in an interactive dashboard.

### Wiseyak Solutions Pvt. Ltd.

**AI,ML Engineer/Backend Developer** *Apr 15,2024 - May 01,2026*

- Built **AI-powered chatbot systems** for **healthcare** and **banking** using **LangChain, LangGraph, FastAPI, and PostgreSQL** enabling **diagnostics**, **appointment scheduling**, and **customer support automation**.
- Developed **SaaS features** including **TTS, ASR, RAG**, and a **secure document-sharing system** with **granular access control**; integrated **pgqueuer** for **background jobs** and **MinIO** for **scalable object storage**.
- Leading **Keycloak integration** for **RBAC-based authentication** and **token workflows**; collaborated on **AI/backend solutions** for clients like **MedAvonix, Siddhartha Bank, Global IME Bank, and Nepal Telecom.**

### Startup AI company: Inspiring lab

**AI/ML/ Data Science Intern** *Dec 04,2023-March 04, 2024*

- Worked on real-world business-oriented AI projects with a focus on resource optimization and future planning.
- Developed predictive models to **estimate future resource requirements** based on population data and demographic trends.
- Gained hands-on experience in **Python**, **Scikit-learn**, and data analysis techniques for **regression** and **classification** problems.

## PROJECTS

**1. Global IME bank ai chatbot (https://chatbot.gibl.com.np/)**
Built backend infrastructure for an AI-powered chatbot using **FastAPI**, enabling real-time query handling and content parsing from the bank's website. Designed and deployed **RESTful APIs** to manage chatbot workflows, automate FAQs, and streamline customer support operations.

**2. SaaS Platform Development**
Contributed to a multi-user **SaaS platform** using **FastAPI**, **PostgreSQL**, and **React**. Developed scalable **REST APIs** with **user authentication**, **role-based access**, and **data management**. Integrated **RAG** and **Generative AI** features to enhance **UX** and automate **content**, while optimizing **database performance** and ensuring secure, real-time **processing**.

**3. Advance Medical Chatbot(Medavonix)**
Developed an **AI-powered medical chatbot** for real-time text and voice interaction, capable of **symptom analysis**, **pre-visit report generation**, and **seamless appointment scheduling** based on doctor availability. Integrated speech recognition and NLP for accurate communication, with real-time database connectivity to ensure up-to-date scheduling and personalized healthcare assistance.

**4. Advanced Role-Based Access System for Document Sharing**
Developed an **advanced RBAC system** for secure document sharing in a multi-user **SaaS platform** using **FastAPI** and **PostgreSQL**. Implemented **granular permissions** based on user **roles**, enabling efficient access management for different document types and user groups. Optimized **database queries** for fast lookups and ensured **secure, real-time access** to sensitive data.

**5. Authentication & Authorization System for Microservices**
Designed and implemented a centralized authentication and authorization system for a **microservices architecture** using **Keycloak**, with **OAuth 2.0 / OpenID Connect**. Enabled token-based, role- and permission-level access control across multiple services, ensuring secure and scalable inter-service communication.