

VIDYA VENKAPPA

☎ +1 206 730 9182 ✉ vivenkap@uw.edu [in LinkedIn](#) [Github](#) [Website](#)

Data Scientist and Software Engineer with 5+ years of experience in data integration, scalable pipelines, and actionable insights. Proficient in Python, SQL, MongoDB, and advanced analytics, with a focus on self-service platforms and data-driven storytelling.

Experience

Cisco Systems

Software Engineer - II

June 2021-June 2024

- Developed a machine learning pipeline that reduced chatbot query response time by 70% by building a retrieval-augmented generation (RAG) pipeline. Managed structured and unstructured data for improved LLM (Large Language Model) responses in the chatbot.
- Ensured responses are polite regardless of the user's query tone, with a 90% positive sentiment score in our custom LLM (Large Language Model) by developing politeness features.
- Achieved a 95% reduction in manual effort by developing a machine learning model that converts flowchart images into pseudocode and translates user queries into SQL responses, seamlessly integrating these functionalities into chatbot systems.

Software Engineer I

Jan 2019-June 2021

- Enhanced bot performance evaluation in BotLite (chatbot framework) by developing an analytics dashboard using Angular's Highcharts and Python libraries, providing visual analytics on user interactions, retention, health status, traffic, and usage.
- Led UI improvements in BotLite (chatbot framework), eliminating redundant features and enhancing user-friendliness, resulting in a spike in bot creations from 400 to 950+ and increase in intent and rule formulations to over 25K+.

Technical Skills

Languages & Tools: Python, SQL, MongoDB, Pandas, Numpy, Matplotlib, Scikit-learn, SciPy, Snowflake, TensorFlow, PyTorch, Spark, Mongo Atlas, Langchain, Azure Copilot, Git, Jira, Miro, PowerBI.

Technical Expertise : Machine Learning, Data Pipelines, Predictive Modeling, Data Query Languages, Data Management, Statistical Analysis, Exploratory Data Analysis, Experiment Design, Generative AI, Big Data.

Projects

Review-Guide: AI Research Paper Evaluator [↗](#)

Dec, 2024

An AI-powered application that leverages Large Language Models (LLMs) to analyze top machine learning conference papers and generate structured evaluations. It extracts key insights, assesses relevance, and provides comprehensive summaries, offering an efficient alternative to human reviewers for academic paper assessments.

Microsoft Digital Defense Report Analysis [↗](#)

Dec, 2024

Analyzed the Microsoft Digital Defense Report (2020-2024) using a Retrieval-Augmented Generation (RAG) pipeline implemented with ChromaDB and LangChain to extract relevant context and identify significant security threats. Leveraged phi-3.5 to generate detailed insights from the retrieved data, delivering actionable findings on emerging risks to enhance strategic decision-making.

Microsoft Security Incident Prediction [↗](#)

Oct, 2024

Developed a machine learning classification model using gradient boosting in Python to analyze security incidents in Microsoft systems, leveraging features from the Guide dataset. Applied regression analysis and feature engineering techniques to optimize the model, achieving an accuracy of 80%.

Spotify Song Popularity Prediction and Recommendation [↗](#)

Oct, 2024

Developed a machine learning pipeline using Spotify's dataset to analyze and predict song popularity. Implemented gradient boosting models to predict popularity based on key audio features such as acousticness and loudness. Designed a personalized recommendation system using K-Means clustering to group similar songs, enhancing user experience with tailored suggestions.

Research Work

Meeting Summarization - SMACC

July, 2023

A system to streamline multi-modal abstractive and extractive summarization.

- The thesis aimed to concisely summarize meetings using Large Language Models (LLMs) for dialogues up to 90 minutes. Introduced a 'Summary Chunking and Chaining' technique for managing extensive transcripts and incorporates an extractive feature to generate video trailers that provide users with quick visual context and key takeaways.

Automatic Subtitle Generation for Videos

March, 2020

IEEE Publications - Conference on Advanced Computing and Communication Systems

- In this project, created a media player capable of generating subtitles for offline videos. Approach involved extracting audio from videos and employing speech recognition engines, including CMU Sphinx and DeepSpeech, to generate these subtitles.

Education

University of Washington, Seattle

Jun 2024 - Jun 2025

Master of Science in Information Systems

In-progress

Courses: Data Analytics, Data Mining, Data Warehouse.

Indian Institute of Science, Bangalore

Aug 2021 - July 2023

M.Tech in Data Science and Business Analytics

3.87/4

Courses: Introduction to Data Science, Data Engineering at Scale, Practical Machine Learning, Deep Learning, Data Mining, Deep Learning for Visual Analysis.

PES University, Bangalore

Aug 2015 - Aug 2019

B.Tech in Computer Science with Data Science Specialization

9.16/10

Courses: Introduction to Data Sciences, Database Management Systems, Big Data, Artificial Intelligence, Machine Learning, Advanced Machine Learning.

Awards and Recognition

- Cisco Emerging Talent of the Year award at ETC India El Premio 2021 (early-career professional).
- First place in the National Cisco Hackfest, earning a project development grant.
- Honored with 25 Connected Recognitions at Cisco for consistently surpassing expectations.
- Secured 2 CNR Rao merit scholarships at PES University, receiving a 25% tuition fee waiver.
- Presented BotLite at various top platforms of Cisco like Cisco Symposium 2021, Cisco Tech Talk, Cisco Connect India 2021 and multiple Webinars.

Volunteering

- Active member of the LGBTQ+ Cisco Pride ERO Core Team, promoting inclusivity and career growth.
- Created chatbots like the "Community Impact Bot" for environmental awareness and also for Cisco's LGBTQ+ community engagement and awareness.
- Volunteered with CareerVillage.org, Concern India Foundation, and GiftAble, developing materials for special needs students and teaching English in government schools.
- Taught a Java class for 100 students, providing individualized mentorship and assessments.
- PR head for ACM Women Chapter, a core member for Aatmatrishya, our college fest and organized events like Samarphana for Defense Forces martyrs at PES University.