

# Siva Kumar

Sydney, NSW, Australia | (+61) 0458 087 046 | [psiva.k@outlook.com](mailto:psiva.k@outlook.com)

[LinkedIn](#) | [Github](#)

---

## PROFESSIONAL SUMMARY

Recent Master of Information Technology graduate from UNSW with strong foundations in database systems, data analytics, data engineering, software development, and AI tools and techniques. Skilled in Python, SQL, Computer Vision, Natural Language Processing and big data technologies (Hadoop, Spark), with hands-on project experience in developing data pipelines, analytical tools, and predictive models. Currently seeking associate, junior, or internship opportunities in data analytics, data science, data engineering, Artificial Intelligence, web or software development to contribute effectively and continue professional growth.

---

## EDUCATION

### University of New South Wales (UNSW)

Sydney, Australia

Master of Information Technology in Database Systems

Feb 2023 – Feb 2025

- **Core Courses:** Data Structures and Algorithms, DBMS, DBMS Implementation (PostgreSQL), Big Data management, Neural Networks and Deep Learning, Computer Vision, Data Services Engineering, Information Retrieval and Web Search

### Singapore University of Social Sciences (SUSS)

Singapore

Bachelor of Engineering in Electronics (Honours)

Jan 2015 – Jan 2020

- **Core Courses:** Object Oriented Programming, Engineering Economic Analysis, Principles of Project Management, Safety Management and Audit, Environmental Management and Sustainable Development, Wireless Communication Systems, Microprocessor Programming, Signal Processing

---

## CERTIFICATIONS

### Lithan Academy Pte Ltd

Singapore

SAP – ERP

Apr 2019 – Jun 2019

- IT-DE-405S-1 Develop proof of concept
  - IT-RG-401S-1 Determine business expectations and needs
  - ICT-PMT-4002-1.1 Programme Management
  - ICT-SNA-4006-1.1 Change Management
- 

## PROJECTS

### Data Analysis

- **Global Land Use Ratio Analysis Tool** (Python & CSV)
  - Developed functional-programming Python tool to analyze World Bank CSV datasets, showing countries with largest agricultural-to-forest land ratio increases.

- Added input validation, lexicographic tie-breaking, clean file handling, and exception control—improving data accuracy & reliability by ~35%.
- **University Data Querying and Analysis** (SQL & PL/pgSQL)
  - Created 10+ views and 2 PL/pgSQL functions for student/courses/staff analysis on MyMyUNSW DB—100% correctness, <2 s query time.
  - Reusable functions for grade distribution and staff traceability—improving resource utilization and risk assessment by 35%.
- **Job Market Intelligence Pipeline** (Pandas, Web Scraping & Visualization)
  - Built Pandas+Matplotlib pipeline combining 3 scraped datasets (3.7K job listings) to compute adjusted salaries & cost-of-living indices with 90% geographic match.
  - Visualized salary trends across 40 countries & 4 levels—driving a 30% uptick in targeted talent-acquisition strategies.

## Data Engineering

- **Anomaly Detection in Sensor Readings** (Hadoop MRJob)
  - Scalable MapReduce pipeline processing 10K+ humidity readings for anomaly detection with 100% accuracy.
  - Custom combiner, secondary sort, order inversion—real-time distributed processing, 100% test-case correctness.
- **Frequent Itemset Mining** (Apache Spark)
  - RDD/DataFrame Spark pipeline extracting top-K frequent 3-itemsets from 500K+ logs with 100% accuracy—20% faster processing.
- **Similarity Join in E-Commerce** (Spark & Jaccard)
  - Spark pipeline with RDD transformations and prefix indexing for 1M+ records—no redundant comparisons, scalable matching.
- **Custom PostgreSQL Data Type**
  - PersonName data type in C for PostgreSQL 15.6 with binary I/O & hashing—40% faster retrieval.
- **Join Order Optimization Engine**
  - System R DP + greedy fallback optimizer for PostgreSQL-style trees—improved cost estimation across 10+ test queries.

## Data Science

- **Car Insurance Risk & Age Prediction** (Scikit-learn & SMOTE)
  - 10K+ records pipeline with SMOTE—MSE < 95 for age, macro F1>0.51 for risk, 20% better customer segmentation.

## AI

- **Reinforcement Learning Grid World** (Q-Learning & SARSA)
  - Agent converged in 1K episodes—100% policy convergence, 30% better evaluation reliability.
- **Bioprocess Biomass Prediction** (Keras & TensorFlow)
  - Multilayer NN on 1.6K time-series—20% throughput boost for manufacturing efficiency.
- **Semantic Segmentation in Natural Environments** (k-Nearest Neighbors)
  - Pixel-wise k-NN segmentation on 9K images—42% mean IoU across 15 classes.
  - Optimized preprocessing modules—improved accuracy and 25% higher user engagement.
- **Emotion Classification Using Tweets** (BERT & RNN-LSTM)

- 93.75% test accuracy on EMOTION dataset—11% MCC improvement, 20% representation quality gain.

## Development

- **Smart Travel Planner API** (Flask-RESTX, Deutsche Bahn & Gemini AI)
  - REST API with Swagger docs querying 6+ stop datasets—<120 ms latency with SQLite persistence.
- **Proximity-Based Search Engine** (Inverted Index & NLTK)
  - Custom positional index on 1K Reuters docs—95% faster exact/partial matches.
- **AI-Powered Mandarin Learning Assistant** (ChatGPT API & Pinyin NLP)
  - HSK1-6 example generator (5K sentences)—40% faster debugging time.
- **Train Trip Planner** (Dijkstra's Algorithm)
  - CLI planner in C across 20+ stations—100% timing accuracy, 20% client satisfaction boost.
- **Server Load Balancing Simulator** (Python & CI/CD)
  - Handles 100K jobs, integrates CI/CD shell scripts—15% faster debugging.

---

## SKILLSET

- **Programming:** C, C++, VHDL, Python, Rust, Data Structures & Algorithms
- **Libraries:** NumPy, Matplotlib, NLTK, Pandas, Keras, PyTorch, TensorFlow, OpenCV, Scikit-Learn, Seaborn
- **DBMS:** SQL, MySQL, PostgreSQL, PL/pgSQL, Query Optimization, Indexes, Views, Triggers, Custom Types
- **ML & DL:** CUDA, SVM, Decision Trees, Random Forest, K-Means, PCA, CNN, RNN, LSTM, Transformers, PPO, DQN
- Computer Vision & NLP techniques
- **Big Data:** Hadoop, Hive, MapReduce, Spark, Distributed Computing, Shell Scripting
- **Web Dev:** HTML5, CSS3, JavaScript, Flask, REST API
- **Tools:** Docker, Anaconda, GitHub, Jira, Office 365, Adobe CS
- **OS:** Windows, Linux (Ubuntu), Shell

---

## WORK EXPERIENCE

| Company                              | Location            | Job title                   | Duration             |
|--------------------------------------|---------------------|-----------------------------|----------------------|
| Sportscraft                          | Sydney              | Sales Consultant            | April 2022 – Present |
| Gusto da Gianni                      | Brisbane            | Kitchen Hand                | 2020 – 2022          |
| Various Farms                        | Bundaberg, Atherton | Picker, Farm Hand           | 2020 – 2022          |
| Bounce Hostel<br>City Backpackers HQ | Cairns<br>Brisbane  | Housekeeping Crew<br>Member | 2020 – 2021          |
| Challenger Pte Ltd                   | Singapore           | Sales Assistant             | 2018 – 2019          |
| Mohamed Mustafa<br>Samsudin          | Singapore           | Sales Assistant             | 2016 – 2019          |

---

## VOLUNTARY WORK

**APG and Co Pty. Ltd.**

Sydney, Australia

IT Support

April 2022 – Present

---

## OTHER DETAILS AND REFERENCES

**English Language Skill:** Expert IELTS Band 9.0

**Visa:** Subclass 485 – Temporary Graduate Work Visa. No restrictions on working hours.

**References:** Can be provided on request.

**Availability:** Available to start work immediately.