

# Jashjeet Singh Madan

▪ San Jose, California, 95134 ▪ 812-803-8144 ▪ [jsmadan@iu.edu](mailto:jsmadan@iu.edu) ▪ [github.com/Jashjeet](https://github.com/Jashjeet) ▪ [linkedin.com/in/jsmadan](https://linkedin.com/in/jsmadan)

## Education

**Master of Science in Data Science | Indiana University | GPA: 3.8/4**

**Aug 2019 - May 2021**

**Bachelor of Engineering in Computer Science | UIT-RGPV, India | GPA: 3.5/4**

**Aug 2012 - Jul 2016**

## Skills

- **Programming Languages:** Python, R, C++, Java
- **Databases:** SQL- Oracle, PostgreSQL, MySQL; NoSQL- MongoDB, SAP HANA, PySpark
- **Course Focus:** Supervised and Unsupervised Machine Learning, Artificial Intelligence, Computer Vision, Natural Language Processing, Data Mining, Deep Learning, Algorithms, Statistics, Data Analysis and Visualization
- **Tools and APIs:** PyTorch, TensorFlow, Scikit-learn, scipy, OpenCV, Spacy, NLTK, SAP BO, Informatica, Oracle Apps, Git
- **Analytical Techniques:** Statistical modeling, Multivariate analysis, Hypothesis testing, ANOVA, A/B testing
- **Machine Learning:** Regression, SVM, Decision Trees, Random Forest, Gradient Boosting, K Nearest Neighbors, Neural Networks, Naïve Bayes, RNN, CNN, LSTM
- **Visualization Tools:** Matplotlib, Seaborn, Tableau, ggplot

## Professional Experience

**Research Assistant | SoCo Lab, Indiana University**

**Oct 2020 - Present**

*Python, Arcgis, plotly, Tableau, pandas, NumPy*

- Collected, cleansed, and provided modeling and analysis of structured and unstructured data of Giraffes mined from social media websites for Smithsonian museum.
- Visualized spatial-temporal data from Cornell's e-bird repository for analyzing migratory patterns.

**Machine Learning Engineer Intern | Nowigence, Inc.**

**Jun 2020 - Aug 2020**

*Python, PyTorch, Natural Language Processing- Spacy, NLTK, Google BERT, ELECTRA*

- Pre-trained the latest ELECTRA and BERT transformer models on a variety of corpus and fine-tuned with transfer learning on GPUs for automatic text comprehension.
- Designed domain-specific multi-label classifiers to classify web pages with an accuracy of 97%.
- Refined Named Entity Recognition models using Spacy with Naïve Bayes probabilistic optimization to categorize unstructured news article tokens from internet with an increased accuracy of 3%.

**Graduate Researcher | Cyberinfrastructure for Network Science Center**

**Mar 2020 - Aug 2020**

*Python, Keras, TensorFlow, Computer Vision- OpenCV, scikit-image, pandas, NumPy, matplotlib*

- Collaborated with researchers from Vanderbilt University to map functional tissue units using computer vision techniques for the HuBMAP project funded by National Institutes of Health (NIH).
- Prototype Object Detection systems for cell identification using U-Net, Alexnet, Faster RCNN, and Mask R-CNN architectures to achieve an F1-score 8% higher than existing models.
- Applied advance data modeling and transformation techniques to derive insights from molecular data by using OpenCV and scikit-image.

**Data Analyst | Tata Consultancy Services**

**Dec 2018 - Jun 2019**

*SQL, Oracle PL/SQL, Python, SAP BO, SAP HANA*

- Identified customer segments using K-Means for RFM (recency, frequency, monetary) analysis, revenue recognition and revenue forecasting.
- Acted as a liaison among a cross-functional team of product owners, business analysts, and technical designers to revamp data reporting which registered 12% increase in YoY revenue.
- Interpreted Key Performing Indicators (KPIs) from large datasets to report financial insights of 46 countries in accordance with International Financial Revenue Standards (IFRS) in Agile methodology.

**Database Developer | Tata Consultancy Services**

**Dec 2016 - Dec 2018**

*SQL, Oracle PL/SQL, Oracle Apps, Informatica, Java, Jenkins*

- Spearheaded the database migration of 20K projects across multiple programs of the biggest IT company to save over \$35M quarterly.
- Guided a team of 3 database developers to automate Invoice Creation APIs as standard PL SQL servers-side routine calls, which reduced turnaround time by 700%.
- Enhanced operational efficiency, and database system response time by 160% by identifying bottlenecks, and query tuning for database optimization using EXPLAIN PLAN and TKPROF.
- Lifted satisfaction score by optimizing and automating ETL pipelines for advanced applications in Sales, Legal, and Finance through Oracle database models and PL SQL objects.