Data Scientist / Data Analyst

ADITYA PRAKASH SINGH

+1 (604) 771-1425 | singh.aditya@alumni.ubc.ca
linkedin.com/in/apsinghanalytics | apsinghanalytics.github.io/
Vancouver, Canada

Work Experience

Strategic Insights Lead / Data Analyst – CadMakers – Burnaby, CA ⠀⠀⠀⠀⠀⠀⠀⠀⠀Jun 2021 – Present
AEC Integration Engineer ⠀⠀⠀⠀⠀⠀⠀⠀⠀Sep 2019 – Jun 2021

- Optimized **SQL**-based ETL processes, utilizing complex queries and **JOIN**s on databases; supported stakeholders (executives, project teams etc.) with critical KPIs, resource scheduling, and analytics through live **Power BI** dashboards.
- Presented complex data in reports to management and stakeholders, driving strategic decisions and operational improvements.
- Conducted data-driven ERP evaluations (Microsoft Dynamics, Oracle NetSuite, Sage Intacct) using cost-benefit analysis and performance metrics to align with technical and business requirements.
- Actively engaged with software development and services teams for feature implementation, focusing on process optimization and enhancing data quality across the organization.

R&D Analyst Intern – Hydra Energy – Delta, CA ⠀⠀⠀⠀⠀⠀⠀⠀⠀Sep 2018 - May 2019

- Formulated a hypothesis and designed experiments; collected and analyzed data using **Python** to identify a substantial amount of unburned hydrogen in truck exhaust, potentially impacting revenue by over $500,000.

Graduate Research Assistant – The University of British Columbia – Vancouver, CA ⠀⠀⠀⠀⠀⠀⠀⠀⠀Sep 2015 - Sep 2018

- Led the systems analysis for a novel clean tech prototype, designing 150+ experiments, and using MATLAB, Python, and advanced statistics (multi-linear regression, ANOVA) to analyze 1M+ rows, enabling industry leadership to make data-driven decisions.

Junior CFD Researcher – Combustion Research Laboratory, IIT Bombay – Mumbai, IN ⠀⠀⠀⠀⠀⠀⠀⠀⠀Aug 2014 - May 2015

- Utilized **Python** with ANSYS FLUENT for high-performance computational fluid dynamics (CFD) time-series simulations on Linux clusters to discover new flame-wall interactions in microscale tubes, leading to 3 publications and URA 02 award (top 0.5%).

Data Science Intern – Indian Space Research Organization – Bengaluru, IN ⠀⠀⠀⠀⠀⠀⠀⠀⠀May 2013 - July 2013

- Built MATLAB algorithms for time-series forecasting of satellite paths, using physics-based models with synthetic noise for generalization.

Projects

CrowdFundProphet App: Crowdfunding Prediction App ⠀⠀⠀⠀⠀⠀⠀⠀⠀**scikit-learn, XGBoost, Streamlit, AWS, Tableau**

- Scraped 24k+ SEC Form C filings using **Python**; curated 7k+ unique campaigns dataset after anomaly detection and removal by statistical methods; visualized trends on a public Tableau Dashboard.
- Boosted model efficiency by reducing features from 27 to 10 using Recursive Feature Elimination, maintaining accuracy.
- Evaluated 20 machine learning algorithms for cross-validation performance, overfitting, and low mutual correlation; selected and tuned stacked classifier (AdaBoost, ExtraTrees, RandomForest, XGBoost), achieving 80% accuracy in crowdfunding predictions.
- Deployed a Streamlit app on AWS that uses the prediction algorithm as backend to assist businesses in optimizing crowdfunding campaigns, via data driven insights on timing, duration, financial metrics, etc.

Education

The University of British Columbia, Vancouver, CA – Master of Applied Science ⠀⠀⠀GPA: 3.9/4.0 ⠀⠀⠀Dec 2018
Major: Mechanical Engineering ⠀⠀⠀⠀⠀⠀⠀⠀⠀*Honors:* GSI Entrance Scholarship, Faculty of Applied Science Award, 1 Publication

Indian Institute of Technology Bombay, Mumbai, IN – Bachelor of Technology ⠀⠀⠀CGPA: 8.5/10 ⠀⠀⠀Aug 2015
Major: Aerospace Engineering ⠀⠀⠀⠀⠀⠀⠀⠀⠀*Honors*: Commonwealth Scholarship, UG Research Award II, 3 Publications

Skills

- **Languages**: Python 3, SQL, DAX, MATLAB ⠀⠀⠀**Libraries:** pandas, numpy, seaborn, matplotlib, scikit-learn, TensorFlow
- **Systems**: AWS, Linux, Git ⠀⠀⠀**Tools:** Power BI (*Certified Power BI Data Analyst*), Excel, Tableau, Jupyter
- **Misc. Technologies:** ETL (MySQL, SQL Server), Regression Analysis, Time Series Analysis, Classification, Hypothesis Testing, Statistics, Data pipeline (mining, wrangling, visualization, modeling, interpretation), Deep Learning, Big Data (Hadoop, Spark)