Satya Kapoor

Burnaby
satya_kapoor@outlook.com

SUMMARY

- 4+ years of experience working in Data Science for e-commerce and product-oriented industry.

- Possess a Master of Science in Economics with specialization in Econometrics by taking advance level courses: Statistical methods in Economics (I & II), Mathematical Methods in Economics (I & II), Intermediate Econometrics, Statistical methods, Advance Econometric Methods, and Applied Econometrics (Panel Data).

- Experienced in handling all types of data: Cross Sectional, Time Series, and Longitudinal/Panel Data.

- Deep understanding of statistical concepts: Correlation, Multi-Collinearity, Heteroscedasticity, Hypothesis Testing, and algorithms: linear/logistic regression, KNNs, random forest, deep learning.

- Experienced in big data: creating data pipelines (ETL) using several AWS resources in PySpark framework.

- Experienced in creating end-to-end projects starting from understanding business problem, converting it into a data problem, onboarding data sets to databases, performing exploratory data analysis and building tradition/ML models, developing web application and hosting models and pipelines.

- Ability to identify root problems by performing deep dive analysis into data and creating relevant tools to resolve the underlying problem.

- Ability to communicate technical concepts and solutions at a level appropriate for technical and non-technical audiences. Conducted webinars for student of Universities (UBC, Langara).

- Excited to learn new technology and analytical methodologies related to cutting edge data science techniques. Regularly participates in bootcamps and trainings of Amazon Machine learning University (Scored Top rank for MLU NLP contest in cohort).

TECHNICAL SKILLS

- Scripting Languages: Python, Visual Basic for Applications (VBA)
- Query language: SQL, Hive
- Databases: AWS Redshift, AWS Athena, MySQL
- Extract, Transform and Load (ETL): Spark, AWS Glue, AWS Lambda, EC2, S3, Sagemaker
- Visualization: MS Excel, PowerBI, QuickSight, Webpages (flask, HTML, CSS)
- Platform: Microsoft Windows, Amazon Linux, UNIX

EDUCATION

- Economics – Master of Science, Madras School of Economics, Chennai, India
- Economics (Honours) – Bachelor of Arts, University of Delhi, New Delhi, India
- Business Administration – Diploma, University of Victoria, Victoria, Canada

WORK EXPERIENCE

Amazon – Vancouver, Canada

December 2021 – Present

Role: Data Scientist

Responsibilities:

- Developing a web application for automating hypothesis testing using Python flask, HTML and CSS on Brazil environment to remove repetitive ad-hoc program/product evaluation task. Created a data pipeline, statistical models, UI integration and full apollo deployment. It helped users to extract on-demand data, and evaluate any program on-demand in few clicks (https://selfservicetool.corp.amazon.com).
- Utilised forecasting techniques to predict the number of customers/learners for an internal e-learning platform using learner interaction and courses data.
- Created several serverless data pipelines using Apache Spark framework, AWS S3, AWS Athena, AWS Lambda, AWS Glue, AWS Redshift and various other resources based on business use case and customer demand.
- Performed topic modeling by working closely with stakeholders for human integration to manually identify the meta topics and mapped the texts/phrases into topics using NLP techniques.

Amazon – Vancouver, Canada

March 2021 – November 2021

Role: Business Intelligence Engineer

Responsibilities:

- Utilised Natural language processing techniques (RAKE, TFIDF, Bert) to extract keywords/key phrases from program participants comments to identify the underlying issues and display it in Word Cloud.
- Built a data pipeline using PySpark, AWS Redshift, AWS S3 and AWS Glue to publish tables containing all metadata and unstructured comments mapped with the participants Id's.
- Performed several hypotheses testing to identify the relationship between the manually generated proficiency and algorithm generated proficiency of the learners.
- Utilised advanced SQL queries and window functions to construct KPI's and views for building dashboard on AWS QuickSight and to meet ad-hoc requests.

Amazon – New Westminster, Canada

Nov 2019 – Feb 2021

Role: ICQA Data and Reporting Analyst

Responsibilities:

- Utilised VBA/Python for accumulating data from various sources like websites by parsing complex JSON, HTML data format.
- Trained statistical learning algorithms to build a predictive tool with high accuracy for identifying Vendor Order ID's with mismatch more than 5% between quantities ordered and received.
- Constructed an interactive dashboard to depict capacity utilization using AWS QuickSight to monitor capacity usages for Canada wide Amazon warehouses.

- Facilitated cross-function departments; by building relevant tools to reduce the processing time, simplified the allocation of labor sharing within the department, and time control of an Associate's
- Automated trackers (Chimebots, Emails) of key metrics assisted in reducing the false damaging of items, it helped in increasing salvage rate from an average of 13% in 2019 to 46% in 2020.
- Built an automated email notification tool to track the associates who work more than 5 hours in the same process, this tool helped in reducing FAIR (First Aid Incident Rate) rate by 18%.

R.A. Malatest & Associates Ltd. – Victoria, Canada

Jan 2019 – Aug 2019

Role: Part-time Project Surveyor

Project description: R.A. Malatest is a research organization, which undertakes government projects to provide quantitative research services to address the current issues in existing policies.

Responsibilities:

- Collected data of multiple sectors such as transportation, education, health, welfare and labor over telephone, emails, and in-person interviews.
- Updated, maintained and prepared raw data for analysis by performing deduplications and missing value imputation.

Edunuts Pvt. Ltd. – New Delhi, India

Aug 2017 – Sep 2018

Role: Research Analyst

Project description: Edunuts is an EdTech start-up. It is a global student community driven platform that assist students in making informed career decisions.

1. To predict the position of leads (potential customers) in the sales funnel.
2. To forecast the market size and narrow down the target market.

Responsibilities:

- Performed Data wrangling and Exploratory Data Analysis for features selection and dealt with issues such as imbalanced data, using over-sampling technique SMOTE (imblearn).
- Trained a classification algorithm to predict the probability of conversion for current leads by using previous leads data in Python (NumPy, Pandas, Urlib2, Scikit-learn, Matplotlib, and Scikit-learn).
- Defined schemas to organise and reduces the administrative complexity of managing separate databases broken down into the logical entities.
- Constructed advance SQL queries to transform and extract data for reporting and ad-hoc queries.
- Developed relevant Key Performance Indicators to track productivity (email engagement rates) and performance of resources (employees, data sources, cost).
- Performed secondary research of the education industry by studying recent policy changes, tracking data on enrollment rates, demographic trends, and drop-out rates.
- My analysis assisted Edunuts to increase leads by 5000 per month and customer acquisition by 30%.

Edunuts Pvt. Ltd. – New Delhi, India

Jul 2014 – May 2015

Role: Junior Research Analyst

Project Description: Build a database from various sources (internal and external) to update the website and generate leads.

Responsibilities:

- Created and maintained database of nationwide universities by discipline, location, and ranking.

- Assisted in developing the 'college predictor tool' for the website by conducting secondary research using a variety of sources (University websites, market research, news articles), increasing total website traffic by 384% in just one year.

- Evaluated data quality testing by ensuring completeness, consistency, uniqueness, and accuracy.

REFERENCES AVAILABLE UPON REQUEST