

NAVJOT SINGH

Contact: (902)943-4733

Email: navjotsingh151@gmail.com

Website: navjotsingh151.github.io



Summary

A versatile and proficient individual with **5 years of experience** in developing predictive models using various machine learning algorithm, natural language processing, providing Business Intelligence Solutions to clients. **Motivated Team player** and organized person with ability to identify customer needs, make technical recommendations and implement effective IT solutions. Seeking a good position as Data Scientist in a progressive organization to leverage the power of analytics to shape better future for everyone

Skills

- **Databases:** MS-SQL, MYSQL, MongoDB, AWS - S3
- **ETL Tools:** Informatica, SSIS, Tableau Prep, Spark
- **Reporting:** Cognos, Tableau, Actuate, Looker, Power BI
- **Programming Language:** Python, R, HTML5, JavaScript
- **Analytic Packages:** Python – pandas, NumPy, scikit learn, keras, OpenCV, nltk, matplotlib etc.
- **Machine Learning Models, Artificial Neural Networks and Web Scrapping, Data Mining, clustering, decision tree learning, regression, A/B Testing**
- **Tools:** PyCharm, Jupyter notebook, eclipse, Spyder, JIRA, Airflow

Experience

DECEMBER 2019 – PRESENT

DATA SCIENTIST, SAINT MARY'S UNIVERSITY- REMOTE

- Working on the **Green Washing Analysis** of 5000+ companies.
- Scrapped 1M+ screenshot using **Selenium and Python**.
- Created predictive model to classify blank and non-blank Images to remove noise from data using **python, keras, TensorFlow, CNN etc.**
- Created Object Detection Model to detect various objects in the screenshot and classify it into feature matrix with 526 different object categories using **python, CNN, AWS S3 etc.**
- Built OCR pipeline to extract text from the 1 Million images using **python, pytesseract**. Cleaned, processed text data and gather the sentiment using **python, gensim, spacy, texblob etc.**

MARCH 2018 – AUGUST 2019

BUSINESS INTELLIGENCE ENGINEER, GERSON LEHRMAN GROUP

- Built **Customer Product segmentation** model which led to **sales** increase by **15% in retail store** using **R, Clustering algo, SQL**
- Built a **Data processing pipeline** to analyze company's KPIs starting from building views/tables in **SQL**, extract and transform in **Tableau Prep** and Dashboard to get insights in **Tableau**. It acts as a **monitoring tool** for the Executive board and Top leadership for measuring company's performance daily.
- Built a **Probabilistic Classification Prediction model** for logistic client to determine the reliability of a port using **Python, Ensemble Algorithm, SQL, NSS Feed**. Exposed the model output through **Rest API using Flask framework**. It has a huge reduction in their **logistic planning phase**.
- Built an **automated testing framework** enabling rigorous data quality check before loading the production data warehouse system for downstream consumption using **Python, TSQL-T, SQL, Apache Airflow**. This reduced the **data quality issue by 25%**.
- Created a Strategy and plan to **build SCD – 1, SCD – 2 and SCD – 3** SQL objects in **Data Warehouse** using facts and dimensions modelling in hybrid (**star & snow-flake schema**) schema and managed source code versioning using **GitHub** along with CI/CD Pipeline. This reduces the **Data warehouse build time by 1.5 hour**.
- Built a **Biography text match tool** to flag the potential profiles to be corrected by Research managers using **MS-SQL, Python, nltk, spacy, fuzzy match, nlp**, following agile project management. It **reduces the research manager profile screening time significantly (30%)**.
- Built **Data collection and Cleaning Pipeline** to extract the data from various sources on **AWS** into **AWS – S3 bucket**. Cleaned data and select features adhering to business logic using **python, Apache Spark** into the SQL Data warehouse.

Achievements:

- Won Special Award in Nova Scotia Open Data Contest- 2020 for improving Healthcare using **Tableau** and predictive modelling in **python**.
- Certified Tableau Developer
- Certified Tableau Analyst
- Certified Tableau Designer
- Certified Tableau Data Scientist
- Member of Corporate Social Responsibility team in the GLG for organizing Charity Events for poor young children.
- Insta Award for the innovation and cost saving for the customer.
- Member of Volleyball team at university and corporate level.
- Part of Khalsa Aid Volunteer team to help poor provide free meals.

Education

SEPTEMBER 2019 - DECEMBER 2020

MASTER'S DEGREE IN COMPUTING AND DATA ANALYTICS

SAINT MARY'S UNIVERSITY - HALIFAX

AUGUST 2011 – MAY 2015

BACHELOR OF TECHNOLOGY (ECE)

PUNJABI UNIVERSITY – INDIA

Experience

JULY 2015 – MARCH 2018

SENIOR SYSTEM ENGINEER, INFOSYS LIMITED

- End to End Infosys TV Analysis, gathering data from SQL and processing and cleaning the data using **Informatica** and provide visual insights from **Tableau dashboard and storyboards**. It enables entertainment and informatics department to switch the content from one segment to another.
- Built **Email Outreach dashboard** for the customer satisfaction department to gather the **statistical insights** on various emails sent to different professional segments using **Tableau, R Server, MS SQL**. It helps to improve outreach channels and **reduced professional churn rate by 10%-12%**
- Built **Automated Access Provisioning Tool** for Reporting support team to service the access request that comes in ticketing tool – Service Now using **python, NLTK, Selenium, SQL scripts**. This tool saves **70 hours** of team monthly time.
- Built **Automation Platform** to decommission unused Cognos reporting tool licenses using **Cognos report studio, SQL Server**. It reduces the company's **licensing cost by 20%**.
- Schedule and maintain monthly and Weekly report delivery of **tableau reports extract using Tableau Server**. It helps sales team leaders to maintain their team targets.
- Managed Multi-platform data collection to merge and analyze large dataset using **spark, python, Data Bricks, AWS - S3**. Collect data in the single place as source of truth for all downstream applications.
- Build, test, and document complex data pipeline which will be help peer to peer knowledge transfer using **GitHub wiki, Visio, lucid charts**.

Project

- Built an **address enrichment solution** in **Service Canada hackathon by classifying address in residential and non-residential** using **Python, AWS - S3, open data, scikit learn, decision tree algorithm**. Solution is selected **among top 5 teams** as it can act as lead generation for investigating team.
- Built award **winning** solution in **NS Open Data - 2020** to build end to end solution from data collection to predictive model. Collection of open data, data cleaning was accomplished using **Tableau Prep**. **Forecasted** the number of patients for diseases using **Tableau**. Built a **multivariate regression predictive model** to determine the staff needed to serve the forecasted patients using **python, stats model, SciPy**. This solution helps the NSHL with better **staff management** and in turn **reduce patient wait time** in hospitals.
- Worked as a **freelancer** for client to gather/scrap all the product categories and price listing from French E-COM website using **python, beautiful soup, request, selenium**. Built a **hierarchy output format to compare different products among various categories**.
- Built the **Storyboard** for Nova Scotia Hospitals to analyze their expenses, overhead, staff allocation using **Tableau, Tableau Prep, python**. It enables Health authorities to better access the **staff needs** and **supply cost** in different hospital
- Created an **Apriori rule-based association mining algorithm** to predict the clicks on website along with **A/B testing** using **R, python, Spark, SQL**.
- Created **web-crawler** for music industry client by data - mining from social media website using **Selenium and python- beautiful soup** and build the **top trending playlist** of songs.