

# Sunny Panchal

## Data Scientist

☎ 6475185427 ✉ sunnypanchal@gmail.com 📍 Ontario, CN in <https://www.linkedin.com/in/sunny-panchal-3b902081/>  
🔗 <https://github.com/28101991SUNNY>

### 📄 Profile

Data Scientist **skilled in Machine Learning and Applied Statistics**, adept at crafting strategic business intelligence solutions. Proficient in leveraging advanced tools like **Databricks, AWS, and Microsoft Fabric** services for impactful insights. Committed to excellence in Data Science and Business Intelligence.

### 🧠 Skills

**Python Libraries and Databases and Version Control** (NumPy, Pandas, Scikit-Learn TensorFlow, Databricks, PySpark, Spark, MongoDB, MySQL, Big-Query, DB2, Oracle, AWS s3 bucket, Github)

**Probability Mathematics and Applied Statistics** (Probability Distributions, A/B test, Confidence Interval Hypothesis Testing: t-Test, ANOVA, Chi-Square Test)

**Machine Learning and Deep Learning** (Supervised Learning ( Decision Trees, RandomForest, GBDT, XGBoost, SVM & techniques including Stacking & voting.) Unsupervised Learning (Time-Series Analysis, Clustering). Artificial Neural Network (ANN) and CNN. Scikit Learn & TensorFlow)

**Data Visualization with Power BI, Tableau & Python** (DAX, Matplotlib and Seaborn, Plotly)

### 📁 Professional Experience

Apr 2023 –  
present

#### Data Scientist

##### Breath of Health Ltd

- Project, Objective and Contributions to Organization:
  - Fostered collaboration to **enhance Early stage colorectal and Breast Cancer Detection for and Covid-19** by analyzing and creating **Machine Learning models on FTIR data**, aligning insights with research goals. Collaborated in the development and preparation of **proof of concept papers** and **presubmission documentation for FDA** approval.
- *Data Science and Engineering:*
  - Orchestrated **PySpark and Databricks** initiatives, boosting data processing efficiency.
  - Engineered end-to-end processes, **reducing preprocessing time by 25%** and enhancing model development speed.
  - Optimized resources with Databricks and EC2 for a 25% to 30% cost reduction.
- *Machine Learning Development:*
  - Developed machine learning algorithms that achieved a **precision of 92%, a recall of 96% and 86% ROC-AUC** using CNNs.
  - Managed **high-dimensional datasets** with **16,666 features** per sample, employing advanced methodologies such as **PCA and t-SNE** for visualization and **clustering**.
  - Utilized diverse **feature engineering** techniques to enhance model performance and resilience, along with experimenting with ensemble methods like **Stacking and voting** classifiers.
- *Statistical Analysis:*
  - Applied rigorous statistical methodologies - **Hypothesis Testing** such as Mann-Whitney, and **t-tests**.
  - Leveraged **descriptive and predictive statistics** for actionable insights and data-driven decisions.
- *Power BI for Data Visualization and Power Platform Apps Usage:*
  - Designed informative Power BI reports and dashboards to meet the needs of stakeholders and executives, facilitating **real-time tracking of clinical trials and sample updates**.
  - **Automated processes** using Databricks and Power BI, leveraging its direct query methods for **seamless integration and data visualization**.
  - Utilized Power BI to analyze and visualize FTIR spectroscopy data, **expediting the analytical process for scientists and FTIR experts**.

Jan 2020 –  
May 2022

#### Quantity Control Engineer


##### Sharp Technocast Industries

- **Monitor, test & report** on the quality of products, Inspect raw materials, components, and final products.
- **Prepared data reports** and business case slides and periodic review reports for internal and external stakeholder review by Utilized BI tools like **Power BI**.
- Developed robust control strategies for complex systems, resulting in **increased reliability and performance**.
- Implemented **predictive maintenance strategies** for control systems, **minimizing equipment failures and production downtime** by up to 25%.
- Performed detailed analysis of process data to identify areas for improvement, leading to optimized plant performance.

---

## Education

---

Jan 2024	<b>MS in Computer Science: AI and Machine Learning</b> <b>Scaler Neovarsity, Woolf University</b> <ul style="list-style-type: none"><li>Pursuing a Master's Degree in Computer Science with specialization in AI &amp; ML through an after-work program tailored for working professionals, allowing for remote study.</li></ul>
Dec 2021 – Apr 2023	<b>Post Graduation Program in Data Science and Machine Learning</b> <b>Scaler</b>  <ul style="list-style-type: none"><li>Attended an intensive program focused on ML and Data Science, gaining practical experience in various aspects of Modeling and data engineering, aimed at upscaling skills to industry standards.</li></ul>
Jun 2009 – Jun 2013	<b>Bachelor of Engineering in Mechatronics</b> <b>Gujarat Technological University, India. (GCET)</b>

---

## Projects

---

### Image classification (fresh produce supply-chain)

- Prepared a MultiClass classifier for identifying vegetables.
- Using TensorFlow, created a Multi Class Classifier with up to 96% test Accuracy, Precision and Recall.
- Used ImageNet dataset in Transfer-Learning, applied base CNN, VGG-19 and ResNet50 architecture.
- The project showcases applicability in various industries, including agriculture, food processing, and retail, with potential adoption by companies such as Whole Foods Market, Monsanto, and Walmart.

### Classifying loan extension eligibility

#### predict the likelihood of loan applicants becoming eligible for an extension or defaulting,

- Developed and deployed classification models that accurately predict with a precision rate of 98% and an f-beta score of 80% . Deployed on Streamlit as well as using Flask, created API endpoint.
- Identified feature importance of crucial parameters for the loan default prediction.
- Utilized ML Models including Logistic Regression, Decision Tree, Random Forest, GBDT, and XGBoost for Prediction, with **applications relevant to leading banks such as CIBC, TD, RBC and Scotiabank, as well as credit reporting agencies like Equifax and TransUnion.**

### E-commerce Business Intelligence Analytics with MySQL

- Analyzed Website Page and Sessions, Conversion Rates across Pages. Done A/B testing on WebPages that are getting more traffic and Products that have higher Sessions Orders Conversion Rates.
- Worked on Bid Optimization, Conversion Funnels, Time Series Analysis - Business Trends and Seasonality, Landing Page Performance, and Testing and Product Analysis.
- The project's insights are applicable across industries, with potential adoption by companies such as Amazon, eBay, and Shopify.

### Ecommerce

#### (Information of 100k orders at E-commerce Company like Target) (SQL)

- Extracted data from Big Query and integrated it with Jupyter Notebook using the Big-Query API. I employed SQL (Big Query) to retrieve the necessary data for business analysis and utilized Python for creating visualizations and conducting statistical analysis on the queried data.
- Performed a time-series analysis of revenue and orders, and calculated confidence intervals of delivery time per state.

### Other Projects:

- Other projects done in Machine-Learning such as Employee Attrition, Telecom Customer Churn Prediction, and Drivers Churn Prediction for ride-sharing companies ,Delivery Time estimation (Regression), Recommendation System for OTT platform, Time-Series Forecasting for an Marketing company

---

## Courses and Certifications

---

### IBM Data Engineering Professional Certificate

### Google Data Analytics Professional Certification

### Microsoft Power BI desktop for Business Intelligence

### Maven Analytics

### Advanced SQL: MySQL Data Analysis & Business Intelligence

### Maven Analytics