

# DATA ANALYSIS ON STUDENT ENROLMENT

---

# Why we are doing this analysis

---

- More than 2.12 million students in post-secondary institutions in Canada
- More than 500, 000 graduates each year
- 42. 77 bn CAD expenditure each year
- Around 30% of the expenditure is spending on Marketing & Sales, which is

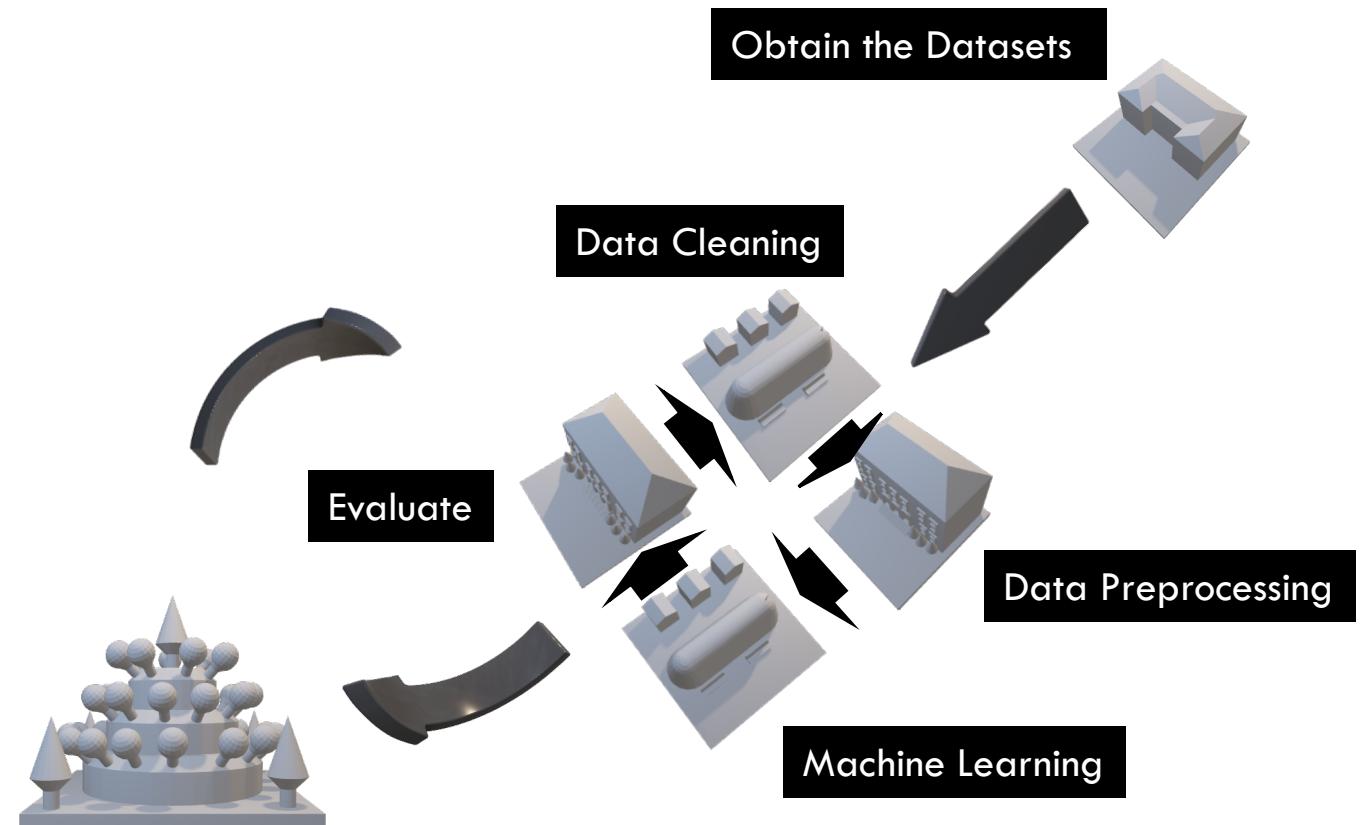


**13 Billion**

# Machine Learning Component

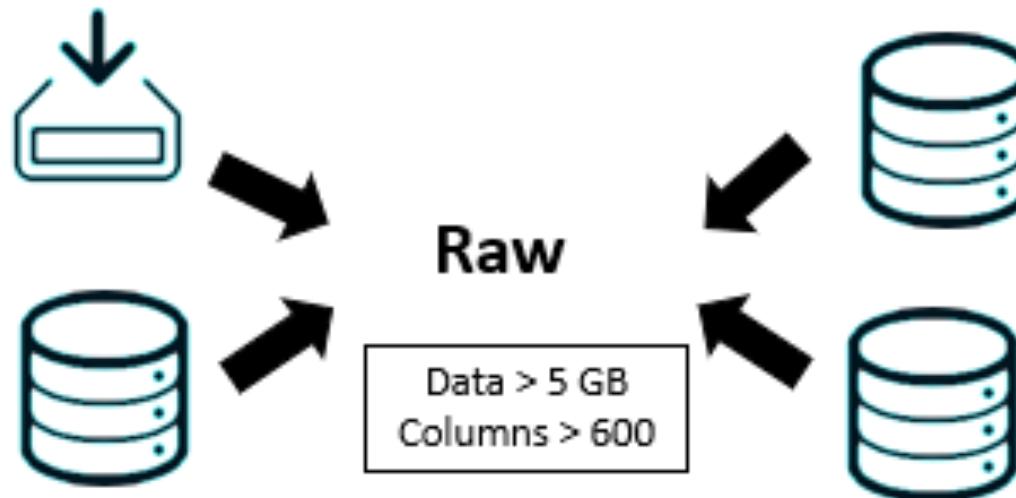
---

1. Obtain the Datasets
2. Identify the problem ( Target )
3. Data Cleaning
4. Pick up the right features
5. Data preprocessing
6. Choose the right models
7. Evaluate the models
8. Try different combinations for better result
9. Finalize



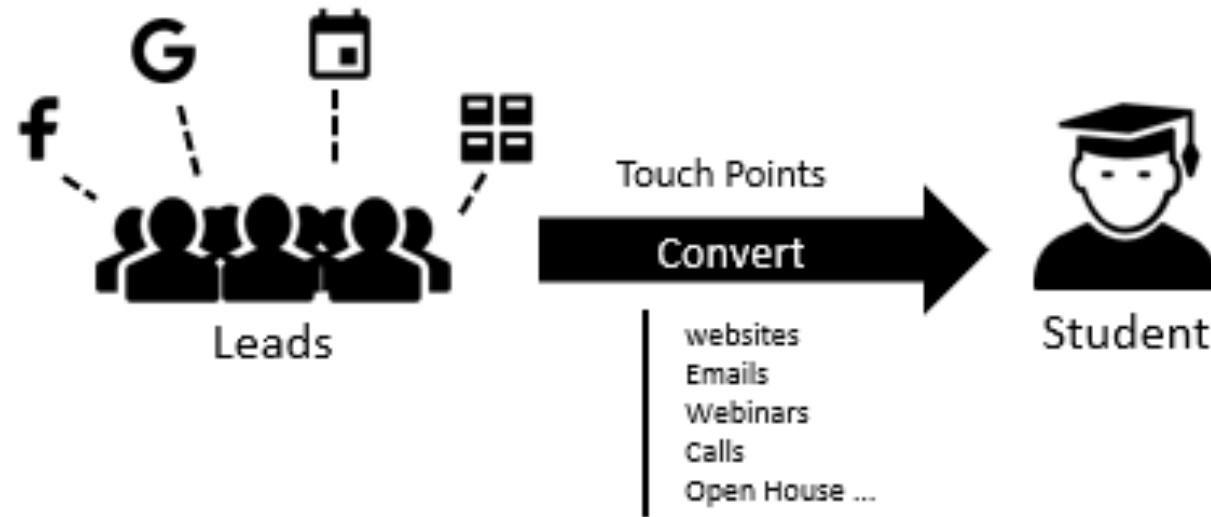
# Obtain the Datasets

---



# Identify the problem ( Target )

---



Based on hundreds of the touchpoints a student interacting with marketing and sales team, how likely a lead is going to convert to a real student

# Process Defined in 4 Easy Steps

---

## 1. Data Cleaning & Pick up the right features

### **Data Cleaning :**

Clean Null ( fill in mean or drop or 0)

Data reformatting ( string, date, float..\_ )

Variance Filter (drop low variance column)

### **Keep import features (try to keep each feature independently)**

Drop unrelated features but still keep essential information

Avoid highly correlated columns !



# Process Defined in 4 Easy Steps

---

## 2. Data preprocessing

MinMaxScaler

Principal Component Analysis (PCA)

StandardScaler

One-hot Encoding

# Process Defined in 4 Easy Steps

---

## 3. Choose Models

Classifier



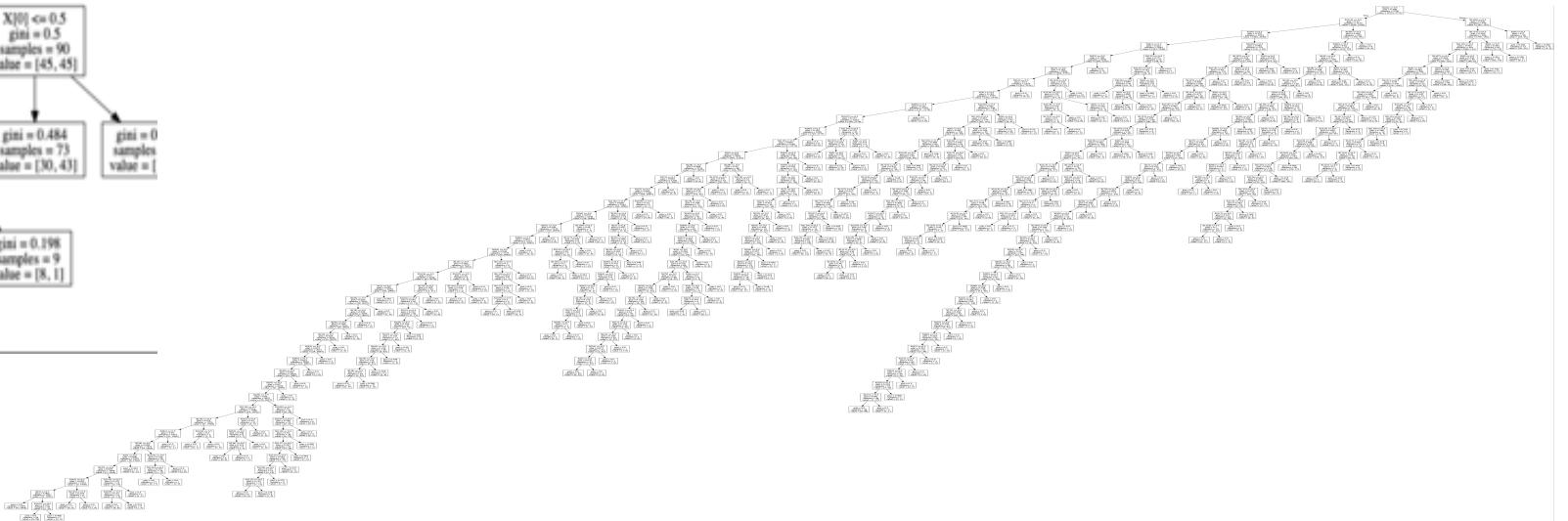
Kneighbors

Decision Tree

Random Forest

# Example Of Decision Tree

---



# Process Defined in 4 Easy Steps

---

4. Evaluate the models and try different combinations

`|GridSearchCV`

Try different parameters and see which are the best parameters

`|Try Different Combinations of data preprocessing methods and features`

# Test Result

---

Kneighbors Best\_Score: 0.85

Best\_params : 'n\_neighbors':10

Decision Tree Score: 0.87



Random Forest Best\_Score: 0.88

Best\_params: 'max\_depth':25, 'n\_estimators':200

# Most important factors affect the conversion

---



INTL Event

Subscribed Emails

Age

Programs

Source

Education Agents

• • •

Personal Score

Website

Country & City

Waive Application Fee

Sales Campaigns

• • •

Not that Important