CASE STUDY FOR PERMIT PREDICTION OF A BUILDING

- Pratima Nagare

Problem Description

- Nowadays, we are living in a concrete world, tall and small buildings here and there, everywhere. But some of them are constructed without any proper permission or documents, while some of the genuine buildings get stuck due to permission delay issues, etc.
- Here's a dataset of building permissions, which consists application date, expiration date, permission granted or not, location, description etc. Based on all these features given you have to predict **which category does the building permission fall into**.

Target/Label

SINGLE FAMILY / DUPLEX

COMMERCIAL

MULTIFAMILY

INSTITUTIONAL

INDUSTRIAL

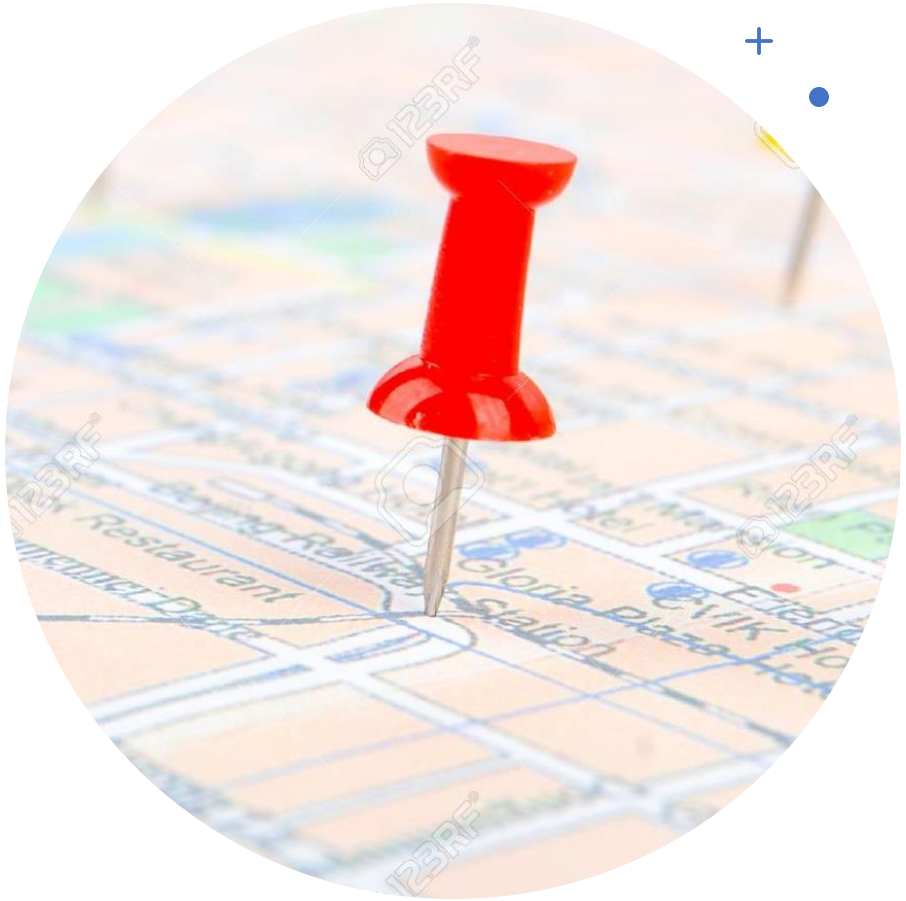
Data Exploration and Processing

High Cardinality Features

- Application/Permit Number
- Applicant Name

Redundant information

- Address
- Location
- Latitude and Longitude



Date columns

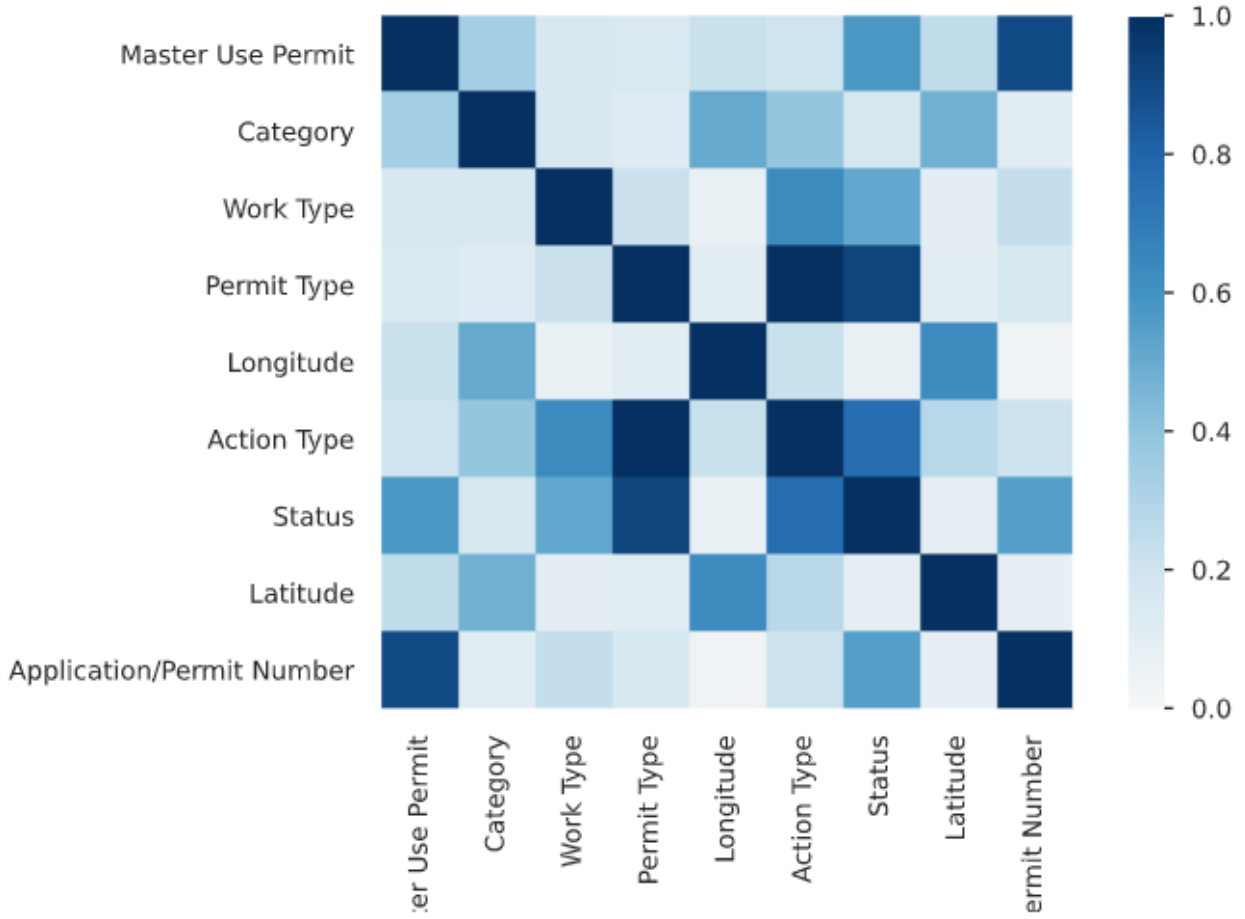
Column Name	Description
Application Date	The date the application was accepted as a complete submittal. If no Application Date exists this generally means the application is in a very early stage.
Issue Date	The date the application was issued as a valid permit. If an Application Date exists but no Issue Date exists, this generally means the application is still under review.
Final Date	The date the permit had all its inspections completed. If an Issue Date exists but no Final Date exists, this generally means the permit is still under inspection.
Expiration Date	The date the application is due to expire. Generally, this is the date by which work is supposed to be completed (baring renewals or further extensions). If no Expiration Date exists, this generally means the application is has not been issued yet.

Date columns

Correlated with status

Effect of time taken for an application on the predication

Correlation



Correlation

Action Type	**Subclassification for type of work being proposed. Valid choices will vary depending on the permit type.**

Combining categories

Status

Value	Count	Frequency (%)
Permit Closed	16407	48.9%
Permit Issued	6052	18.0%
Initial Information Collected	5185	15.5%
Application Accepted	1522	4.5%
CANCELLED	1425	4.2%
Reviews Completed	540	1.6%
AP Closed	478	1.4%
Permit Finaled	240	0.7%
Information Collected	60	0.2%
Cert of Occupancy Authorized	51	0.2%
Other values (2)	5	< 0.1%
(Missing)	1574	4.7%

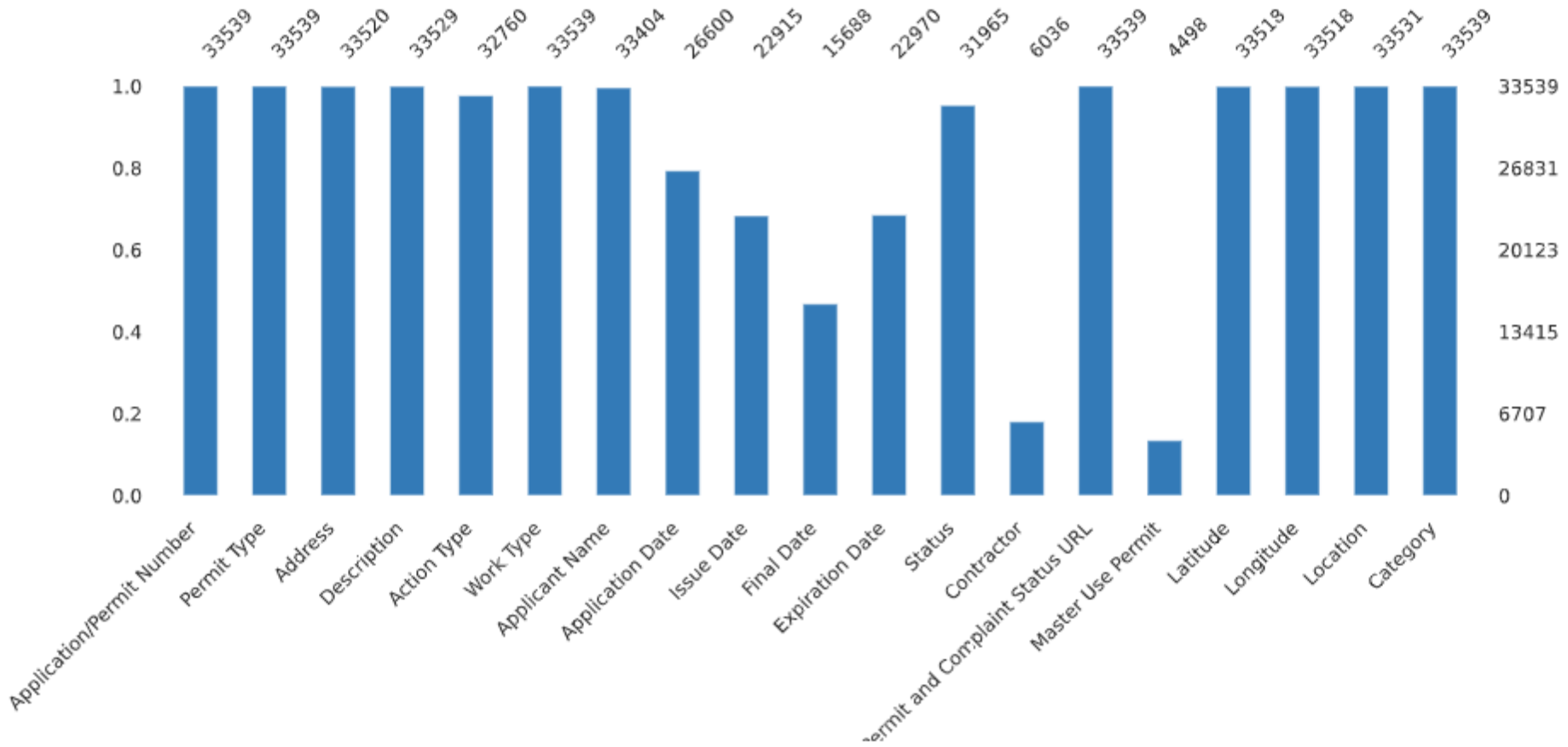
Action Type

Value	Count	Frequency (%)
ADD/ALT	21382	63.8%
NEW	6864	20.5%
DEMOLITION	2593	7.7%
ALTER	933	2.8%
TREE/VEGETATION MAINT/RESTORE	408	1.2%
GRADING	194	0.6%
TEMP	170	0.5%
NO CONSTRUCTION	110	0.3%
CURB CUT	74	0.2%
SHORELINE EXEMPTION ONLY	10	< 0.1%
Other values (6)	22	0.1%
(Missing)	779	2.3%

No Direct Information Gain

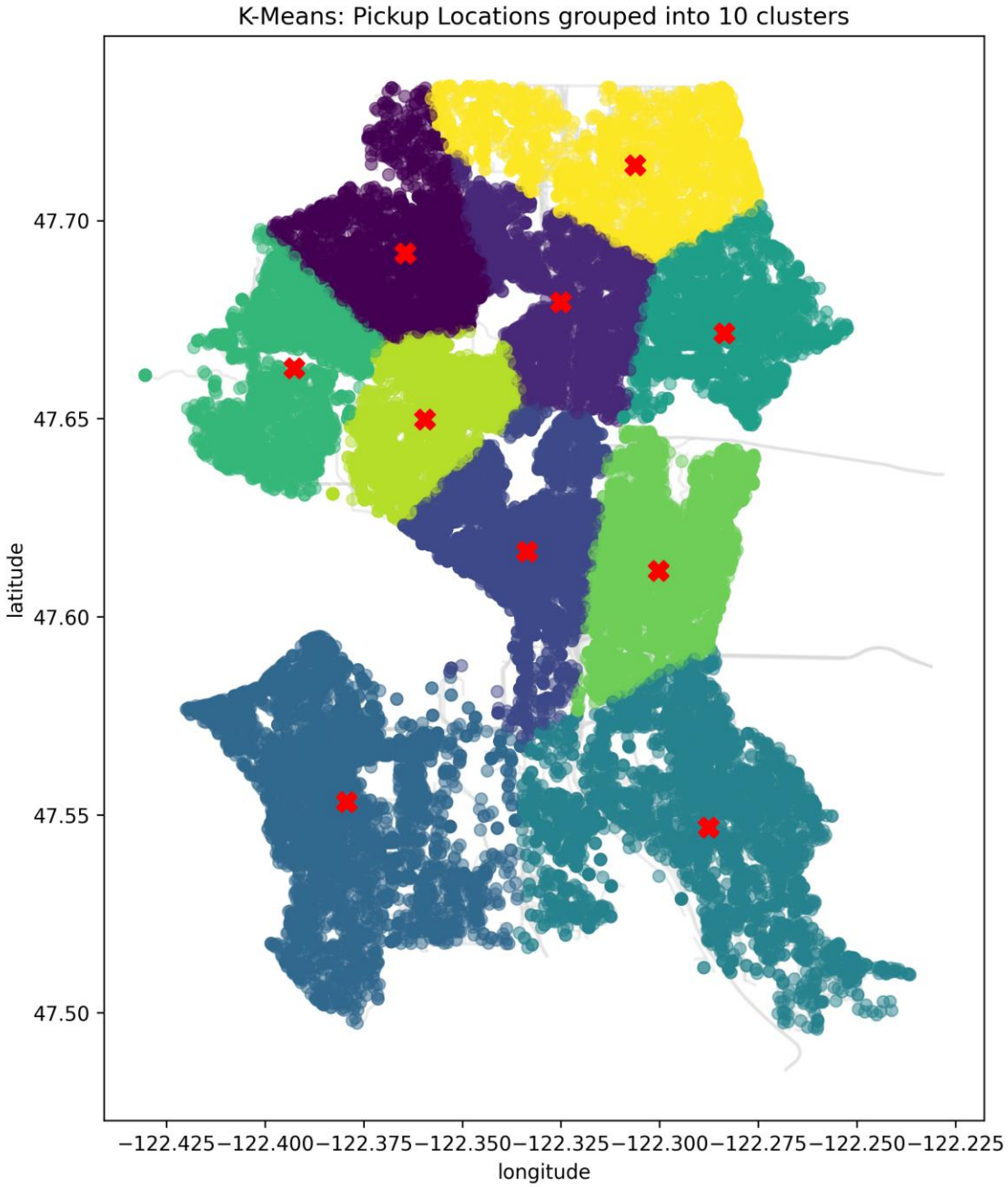
- **Permit and Complaint Status** - Link to view full details and current status information about this permit at Seattle DCI's website.
- **Description** - Brief description of the work that will be done under this permit.

Missing data

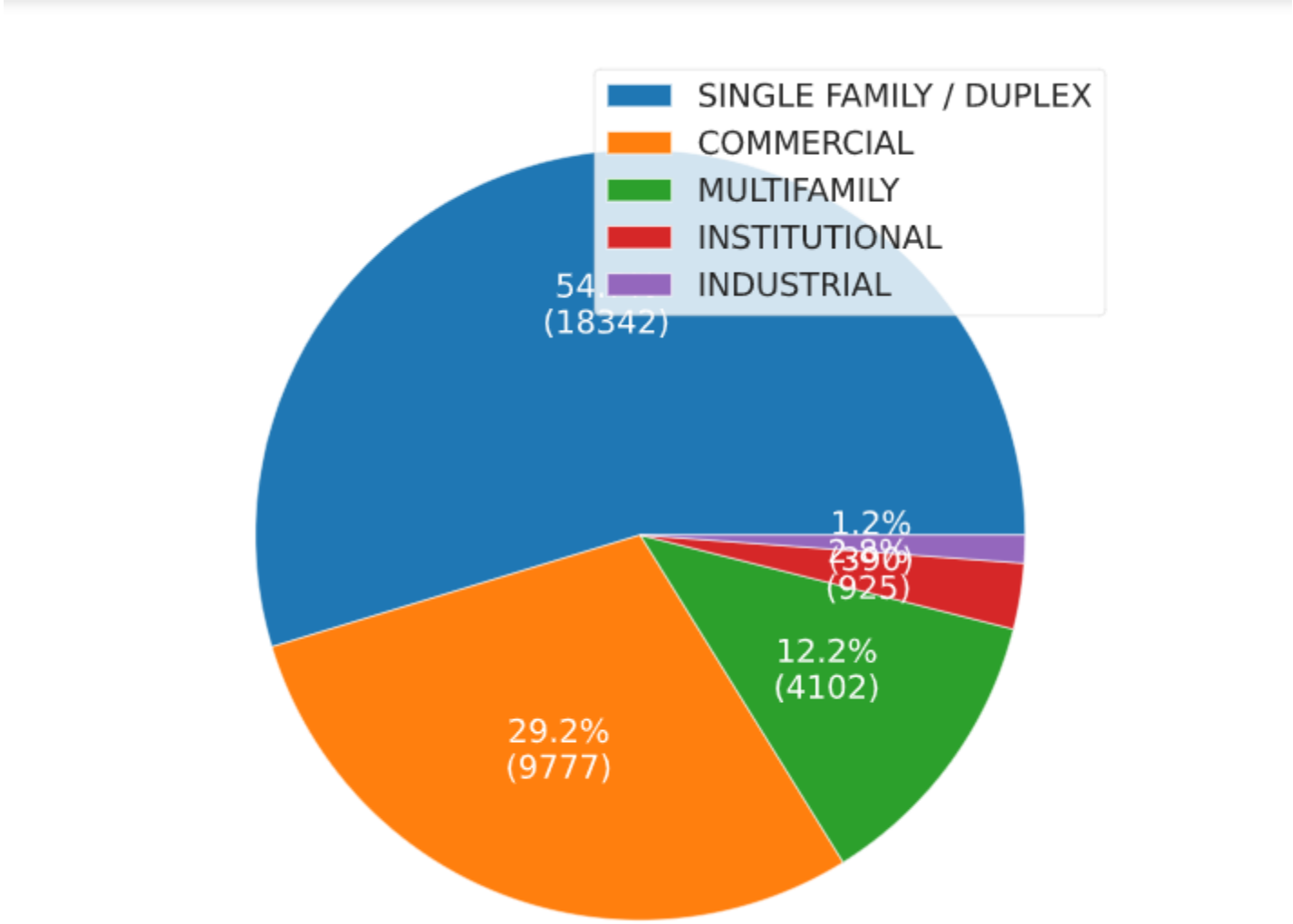


Geospatial data

- Reverse Geocoding - recover address from geographical location
- Clustering



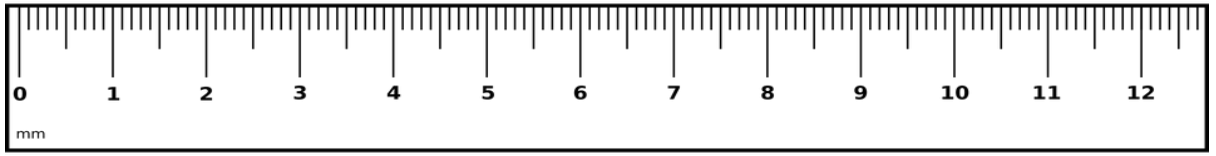
Target Variable Distribution



Encoding Categorical Features

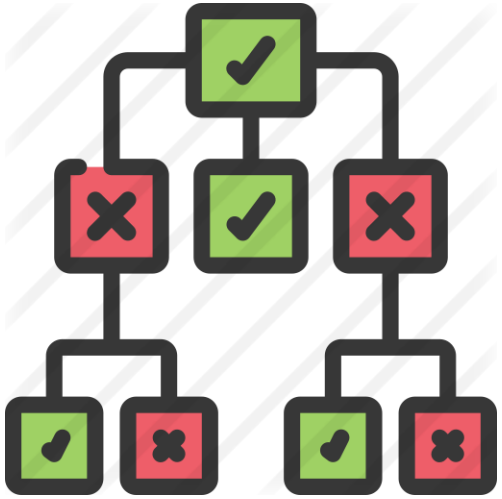
Train Test Split

Weighted f1 score

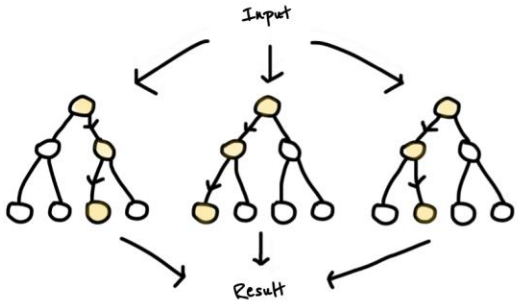


Metrics

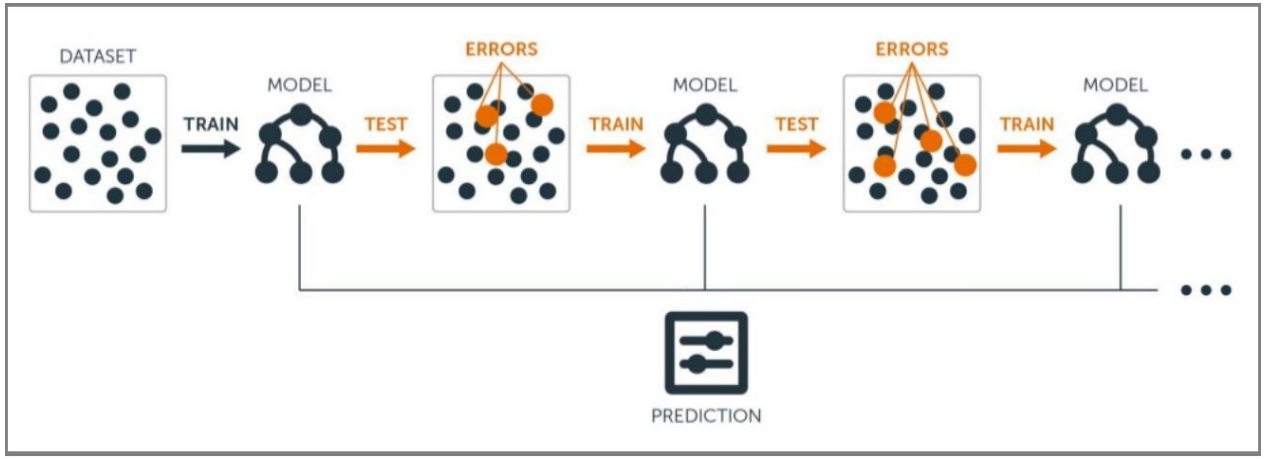
Models trained on data



Decision Trees



Random Forest



Gradient Boosting(AutoML)

Some more options to try

Description

Models

Hyperparameter tuning

Any suggestions?