

Eg: Control center Dashboard

① Understanding the business problem

→ Classification of incidents from news feed (high volume)  
Cot of man power required -

② Feasibility testing.

→ Has this problem been solved with ML in past.

→ Literature Review → OK

③ Defining success metrics → based on HLP or SOTA → 95%.

Classification accurately → 95% or 5% error rate.  
precision  
recall/f1

← news feeds  
← news feeds  
manually classifying  
→ irrelevant  
→ natural emergency  
→ traffic incident  
→ social unrest  
→ Criminal activity  
Accenture facilities

④ Identifying data requirements

→ Training data → Already classified news feed incidents / supervised

⑤ Collecting training data

→ Proprietary customer data ✓  
→ Open source data sets ✓  
→ Web Scraping ✓  
→ manually labelled ✓

Historically labelled data  
+  
Training data

→ Cleaning and preprocessing data

numerical data  
→ Imputing missing values  
→ Fixing data imbalance

→ Normalizing data  
→ Dimensionality Reduction

PCA TSNE

→ Train Test split

→ removing empty txt examples  
→ irrelevant: 90% remaining all classes: 10%

10,000 → 9000 - irr  
ROSE - Random over sampling of minority data

2000 ← 100 - Acc  
2000 ← 200 - Acc  
2000 ← 300 - crime  
2000 ← 800 - Social unrest

→ Creating a baseline model

50k → training data → train 80%  
validation - 20%

(40k) baseline → 80% accuracy = 20% error →  
(10k)

2k error records → error analysis

40k

(10k)

2k error records → error analysis

→ One particular class has low accuracy.  
↓

→ Improving model performance (based on error analysis & data driven)

→ √98% acc || pre-trained model → fine-tune.

→ Model deployment  
→ Monitoring model o/p for drift.

→ best model → Wikipedia corpus for language modelling.  
[predicting the next word]

Final prod. model

@ deployment

24 hours news feed

80% irr

→ ±10%

10% traffic

→ ±10%

→ alert

5% crime

3% disaster

etc.

① Understand the business use case:-

Figure out the client requirement → eg:

Automate Disaster mgmt team's manual classification of news feed in different categories eg, irr, crime, etc.

② Figure out the feasibility:-

Literature Review

③ Success metric based on literature review results  
✓ SOTA models  
✓ AUP

→ Classification

→ Acc. → 98%

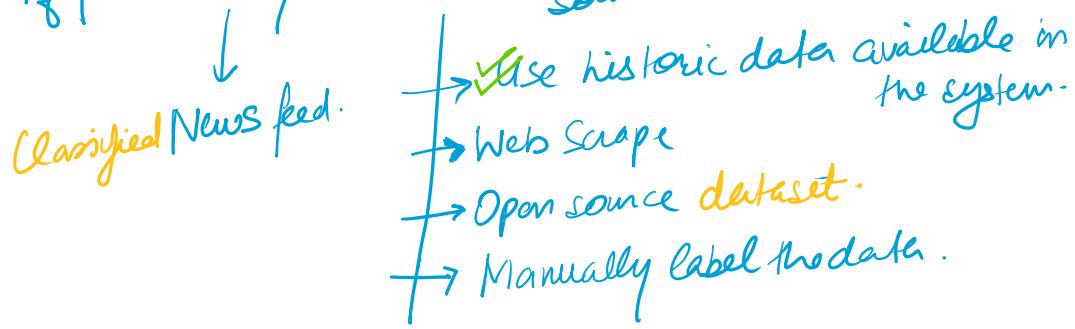
→ Recall

→ Precision

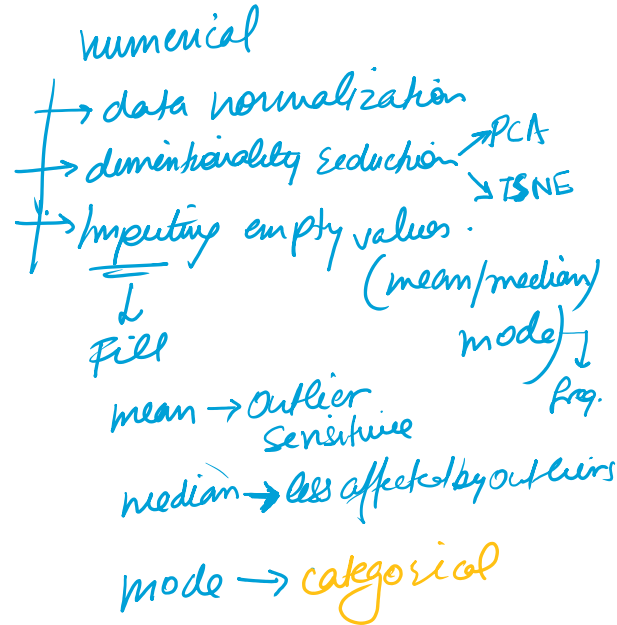
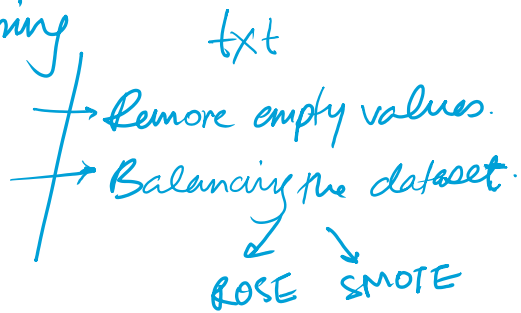
④ Identify data requirement & data sources.

data available in

④ Identify data requirements & sources.



⑥ Data Cleaning



⑦ Baseline Model.

→ 80% acc

20% error → error analysis in data & model.

⑧ Model Improvements

After the success metric is met.

⑨ Model Deployment

⑩ Model maintenance → Output distribution of labels.  
monitor karte hain.