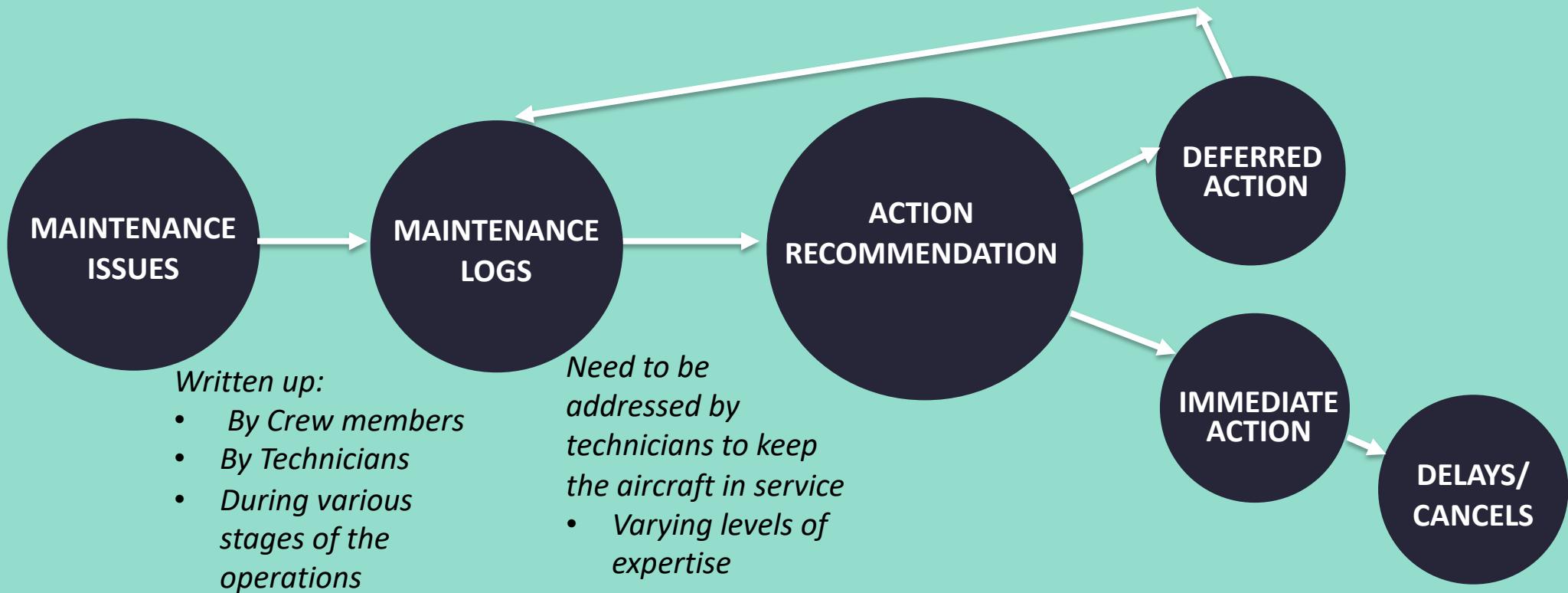




NLP Modeling for Aircraft Maintenance

Deepa Mahidhara
May 17, 2019

Context



Process for troubleshooting maintenance logs is very manual, time-consuming, and prone to human errors.

Problem Statement

Build a Machine Learning model to:

- Read maintenance logs
- Convert them to machine-readable language using NLP techniques
- Predict an action based on historical data

Benefits

- Speed up the maintenance process
- Reduce human errors
- Reduce airline delays & cancellations

Types of Maintenance Issues

ATA Chapter	ATA Description
ATA 09	TOWING AND TAXIING
ATA 10	PARKING, MOORING, STORAGE AND RETURN TO SERVICE
ATA 24	ELECTRICAL POWER
ATA 25	EQUIPMENT /FURNISHINGS
ATA 48	IN FLIGHT FUEL DISPENSING
ATA 57	WINGS
ATA 61	PROPELLERS/ PROPULSORS
ATA 76	ENGINE CONTROLS
ATA 79	ENGINE OIL
ATA 92	ELECTRICAL SYSTEM INSTALLATION

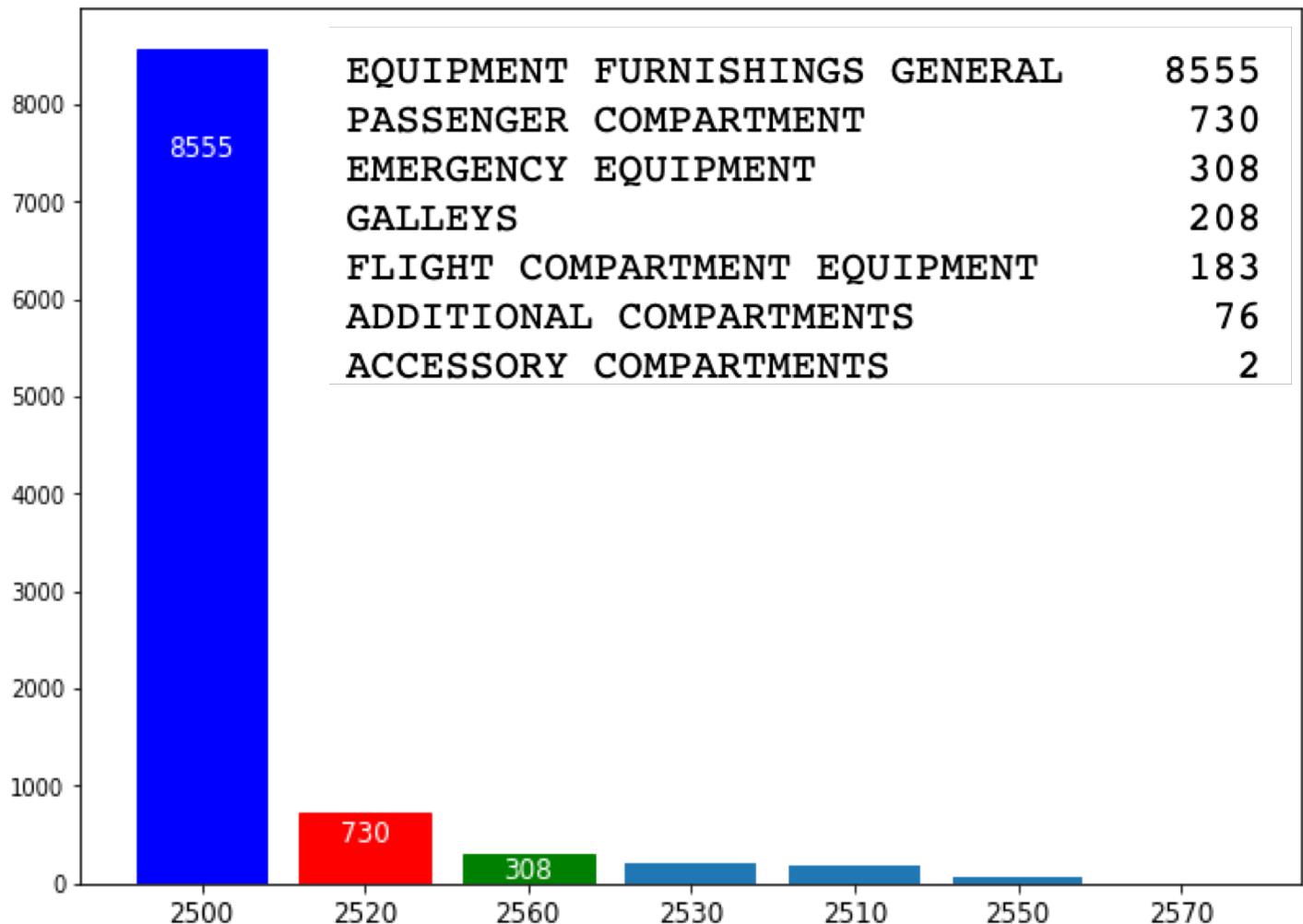
For this project, we are looking only at ATA25 codes, which deal with equipment & furnishings inside the plane.

Example: Original Text

problem_log	corrective_action	ata4_desc
<p>CREW REPORTED RUBBER SMELL IN CABIN FROM ROWS 1 TO ROWS 21. NEWMAN TOMC. SAMC TEXT: WO#15092916. CREW ALSO REPORTED THAT SMELL HAD QUICKLY DISAPATED. NEWMAN TOMC.</p>	<p>:NO DEFECTS NOTED IN GALLEY EQUIP, CABIN EQUIP, AEVC EQUIP AND AIR COND EQUIP PER TSM 05-51-00. FO ALLOWED TSM 21-00-00-810-801 PERFORMED ENG AND APU RUNS AND NO BURNT RUBBER SMELL PRESENT. NEWMAN TOMC. - OPBLM NEWMAN JAY</p>	PASSENGER COMPARTMENT
<p>LOUD CHATTER AND VIBRATION HEARD AND FELT IN AC.</p>	<p>:RESET VALVE IN WATER LINE BEHIND COFFEE MAKER AFT GALLEY. NOISE STOPPED.</p>	GALLEYS
<p>29DEC MFE DF-Q :AFT AFT LAV TRASH DOOR FLAPPER SPRING IS BROKEN. NEWMAN TOMC. SAMC TEXT: WO# 15122914.</p>	<p>:DEFERRED PER MEL. TAPE OVER FLAP PER DOOR AND INSTALLED DF-CF PLACARD ON FLAPPER AND LOCKED OUT LAV DOOR. NEWMAN TOMC.</p>	EQUIPMENT FURNISHINGS GENERAL

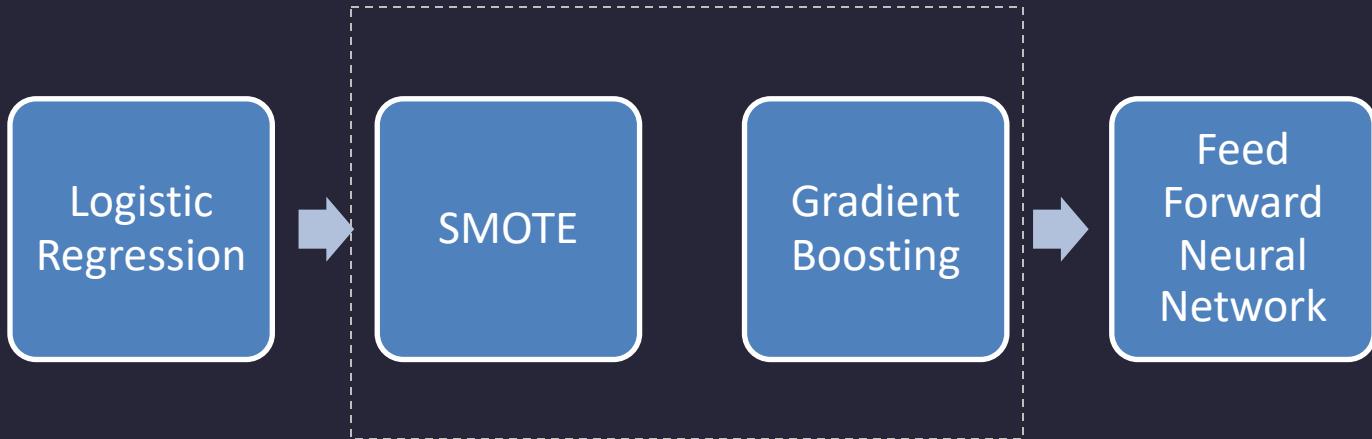
Airline has classified logs into broad descriptions using historical relationship between logs & actions; objective here is to build a model to predict action.

Highly Imbalanced Classes



Predicting classes is challenging as Class 2500 is 85% of the observations.

Modeling: Predicting ATA Classes



Results of FFNN:
Training Accuracy: 0.85
Test Accuracy: 0.84



Accuracy of the models was quite good, and there was no overfitting problem ...



... however, the models did not address the basic business problem, i.e.

- predicting corrective action

Problem Log

Corrective Action

Using Cosine Similarity

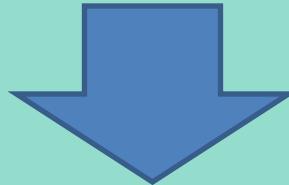
PROBLEM LOG: [coffee maker 812 flowing constantly]

problem_log	cos_similarity	corrective_action
coffee maker 812 flowing constantly	1.000000	removed and replaced coffee maker 812 ops and leak check ok rotatable parts change was indicated tt 1431005 sn off 06210coffeemaker a off 78 2500 9 9001 6210 on 78 2500 9 9001 9805
fwd coffee maker will not stop brewing water is constantly coming out	0.853548	removed and replaced coffeemaker at position 212ops check normalrotatable parts change was indicated tt 1465458 sn off 0688coffeemaker in off 31 2534 9 0001 0688 on 31 2534 9 0001 0010
coffee maker 509 mid galley constantly overflows water	0.841361	rotatable parts change was indicated tt 1673152 sn off 07610coffeemaker a off 78 2500 9 9001 7610 on 78 2500 9 9001 3907

Identify historical log with highest cosine similarity and find corresponding action.

Using Spacy

PROBLEM LOG	NOUNS and VERBS
coffee maker 812 flowing constantly	Noun phrases: ['coffee maker'] Verbs: ['flow']



CORRECTIVE ACTION	NOUNS and VERBS
removed and replaced coffee maker 812 ops and leak check ok rotatable parts change was indicated tt 1431005 sn off	Noun phrases: ['coffee maker'] Verbs: ['remove', 'replace', 'be', 'indicate']

Next Steps

- Data needs more cleaning to isolate date, other numbers from valid numerical data
- Try to find a corpus of tech words
- Explore Spacy's Topic Modeling to:
 - Better model time & other numerical data
 - Discover specific topics in historical data
- Use Unsupervised Learning to model these relationships