

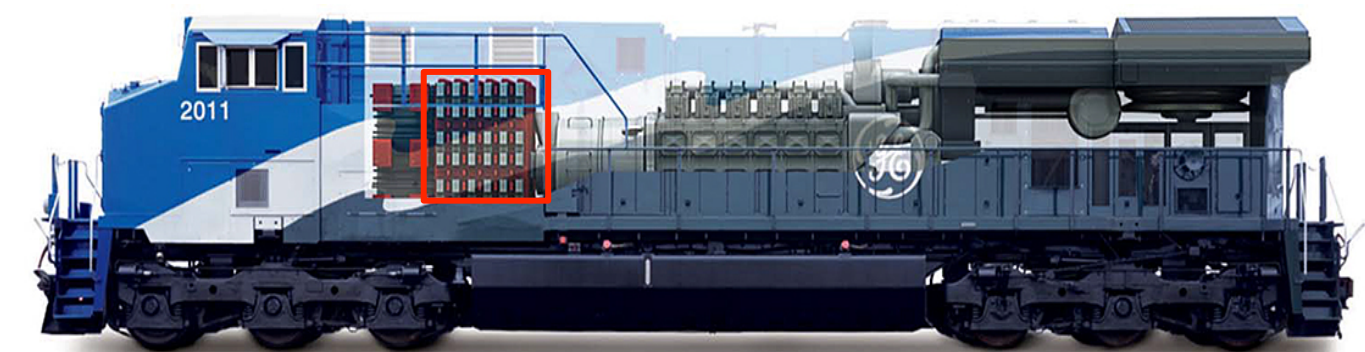
Situation & Complication

- Maintenance services of locomotive have been based on “standard, plain-vanilla” schedule for over a CENTURY.
- If data analytics make condition-based maintenance possible, significant cost reductions and performance improvement can be achieved
- Inverters convert DC to AC to power motors. When they break, motor stops
- Replacement activities generate text data. Comments from Mechanics

Research Questions

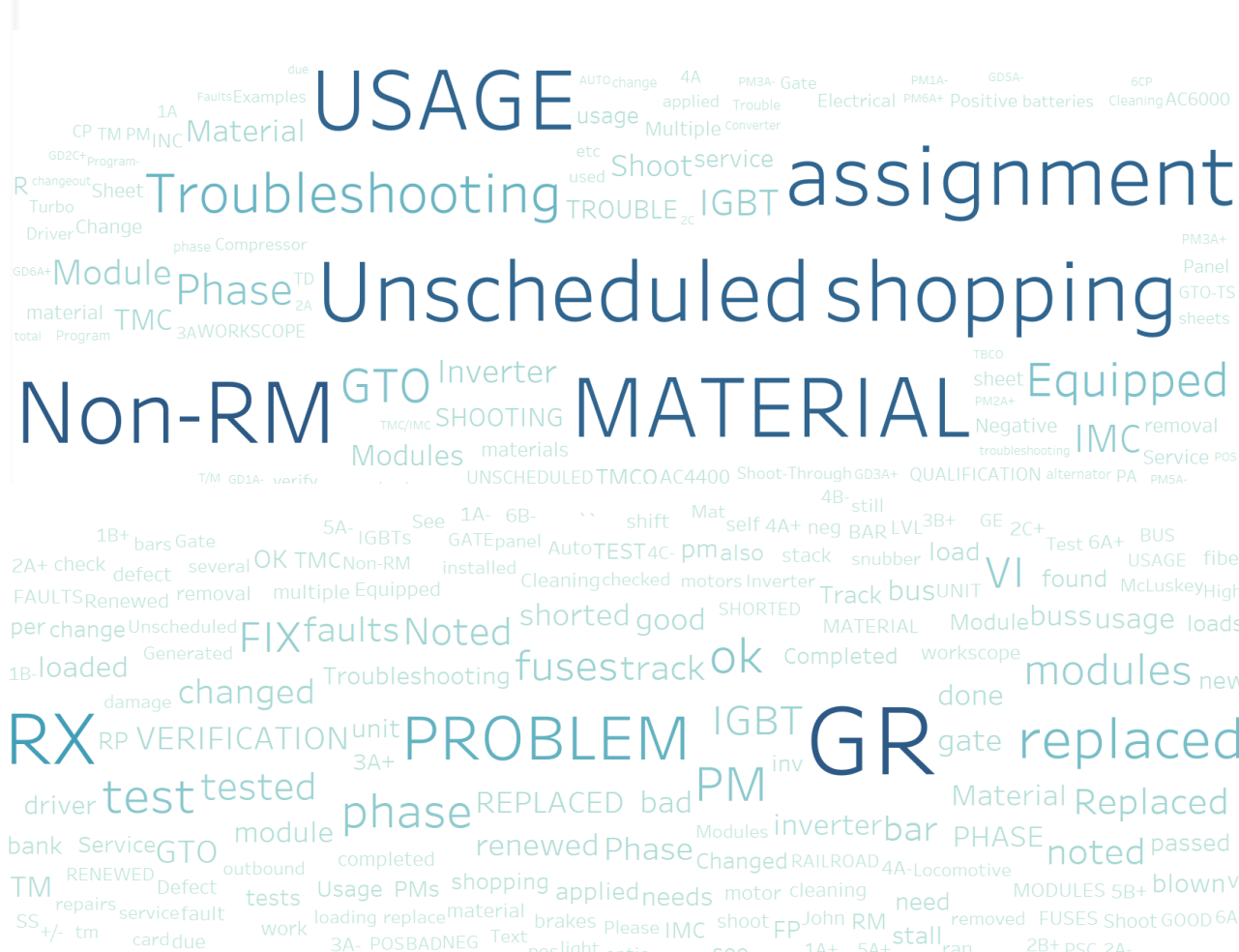
Can free-text data help predict early failures after replacement?

- *Does free-text complement other service data?*
- *What types of comments/descriptions are most useful for predictions?*



Data & Feature Generation

Word Cloud of Data Set



Features generation involved multiple text-analytics methods



Word (Stem) Parsing

Bag of words for most common terms

Free-text processing

Typo/Punctuation/Stop words removal + Word stemming

Syntax Analysis

Occurrence of verbs, nouns
and adjectives

K-means clustering

Grouping of key terms
based on Euclidian distance

Sentiment Analysis

Sentiment score for text in each field
– syuzhet algorithm

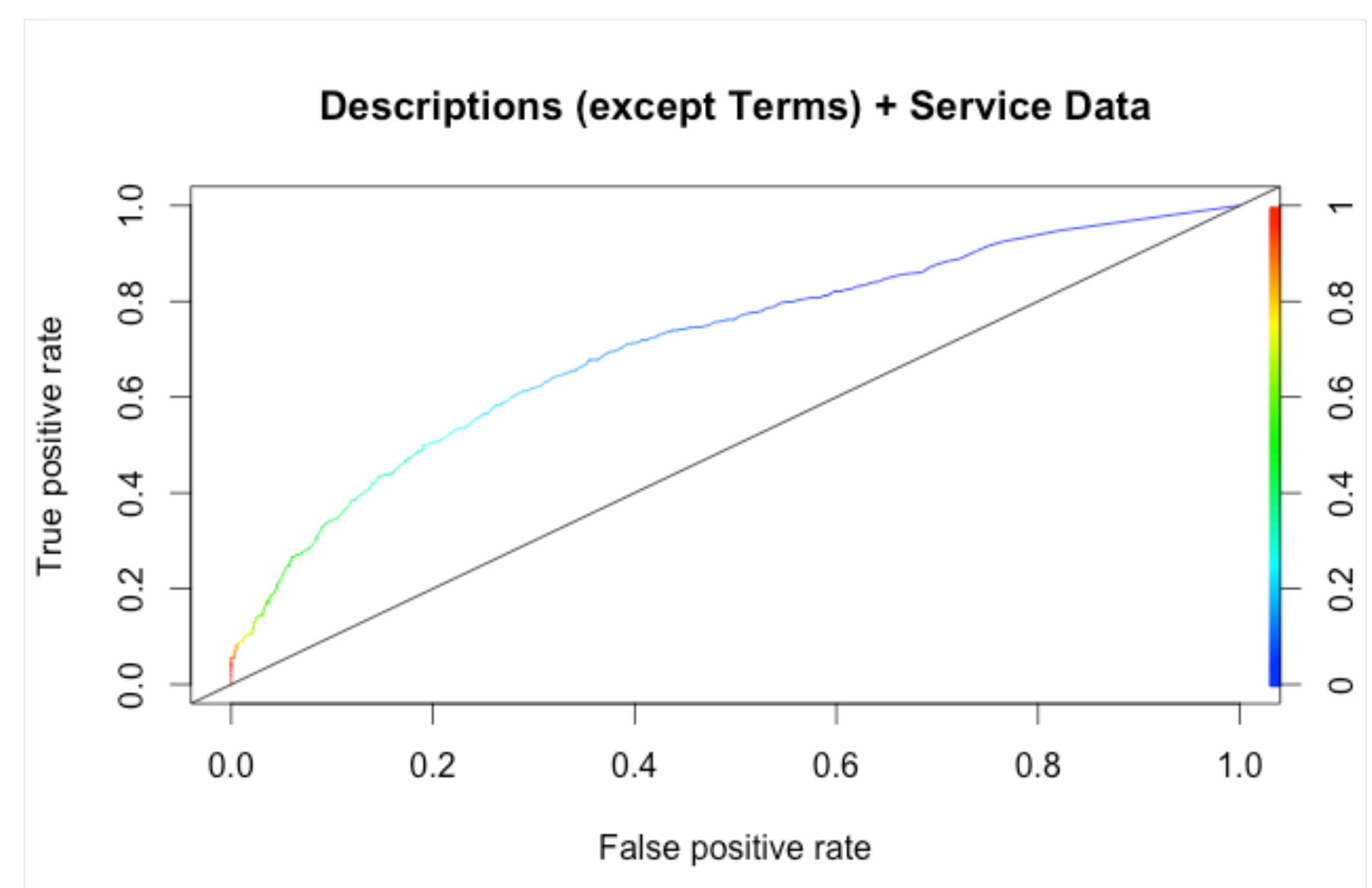
“277 Features” in Total

	Feature Group	No. of features
Text Structure	Length	2
	Sentence syntax	6
Text Content	Sentiment	2
	Word Clusters	6
	Term occurrence	253
Service data (Part type, Position, Service type)		8

Prediction Models & Performance

Feature Group		All text features	Reduced features	Reduced features + Service data
Structure	Length	Comment + Description	Descriptions	Descriptions
	Syntax	X	X	X
Content	Sentiment	X	X	X
	Clusters	X	X	X
	Terms	X		
Service data				X
AUC		0.693	0.675	0.711
Accuracy		0.796	0.797	0.786
No. of features		270	9	16

Random Forest model generates AUC of 0.711 with 16 features



Insights & Learnings

- Free-text **descriptions of replacements can predict early failures** (.675 AUC, 9 features)
 - Ensemble algorithm (Random Forest) performs significantly better than regression
 - Best model combined features for both content and structure of the text
- Text-data complements Service Data**, improving performance (0.771 AUC, 16 features)
 - Further combination with sensor data may help improve current models, informing maintenance decisions
- Very **Short/Blank descriptions are not very good indicators**
 - By reinforcing need for documentation, it is possible that GE can train even better model