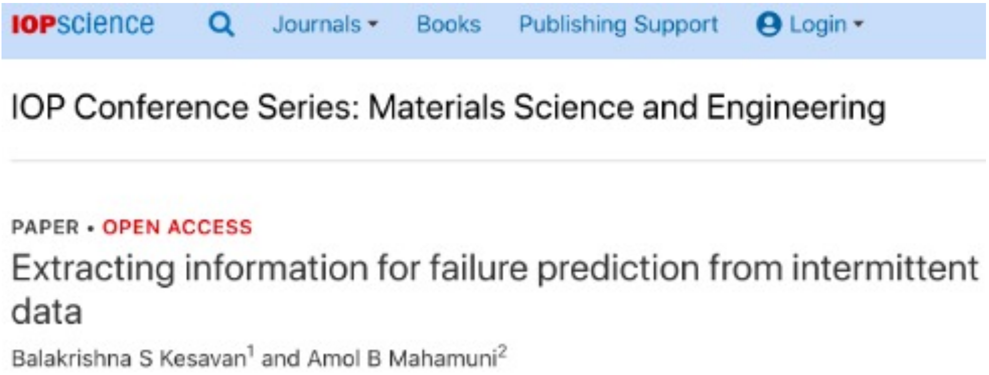
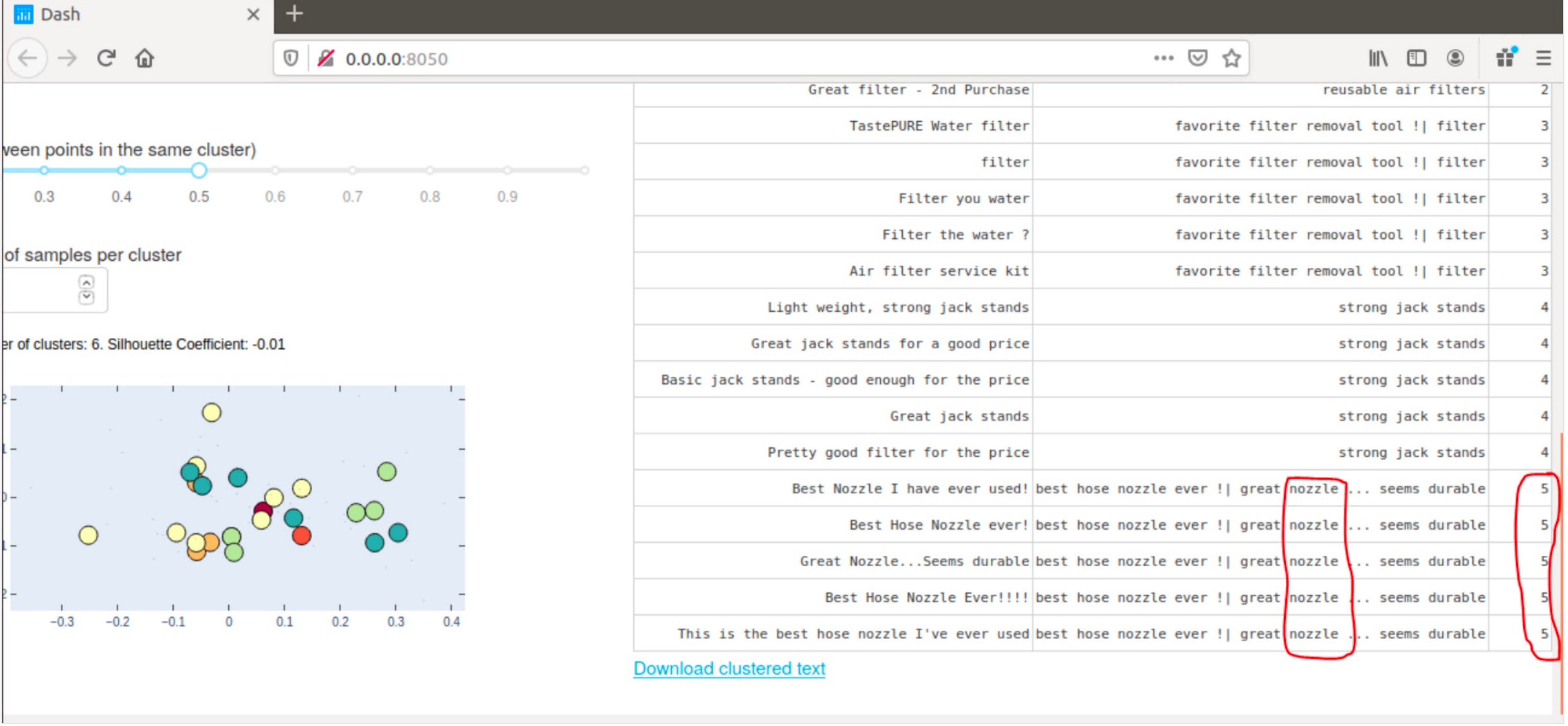
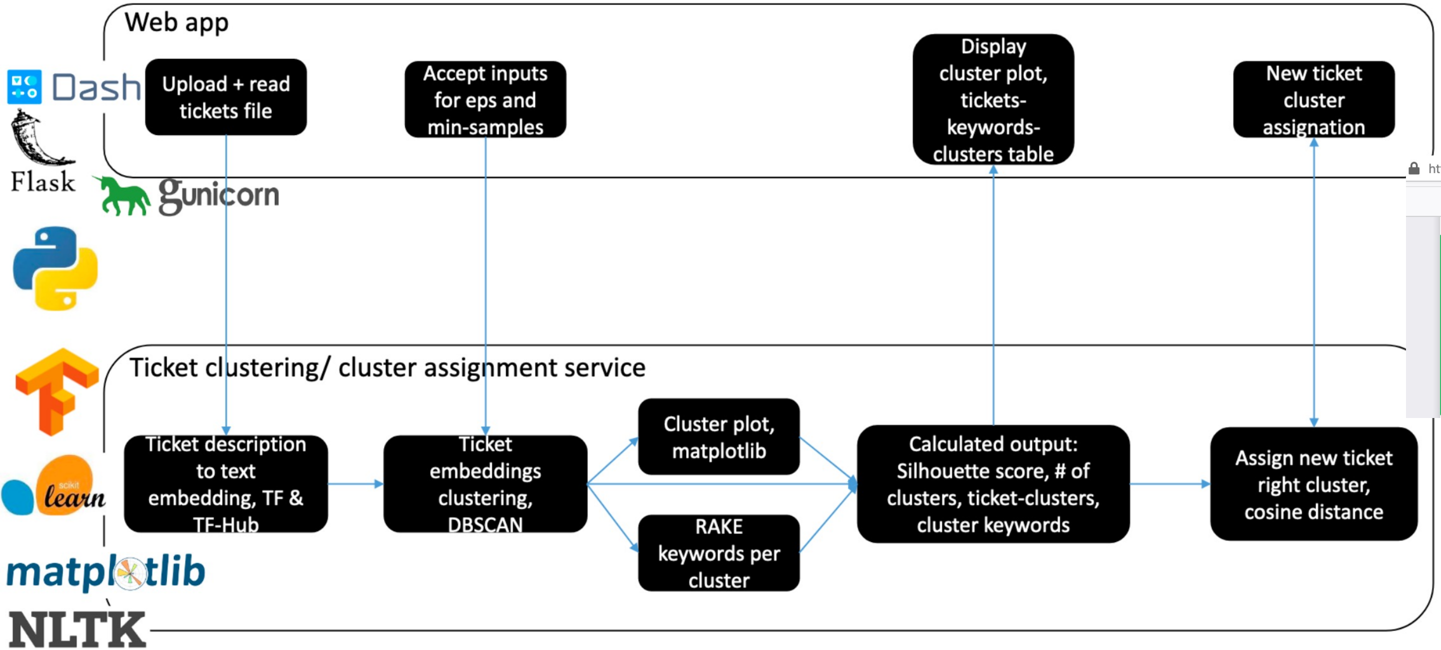
Overview

Data Science projects since 2014: Pharma, Utility, Predictive Maintenance, Services
10 years US onsite: business process, ERP/BPM, management

	Tabular data	Text
ML	1. Reducing call center load (ARIMA) 2. Failure prediction (survival) 3. Optimizing project staffing (RF)	1. Measuring project health (polarity & valence) 2. Azure helpdesk chatbot (cloud)
DL	1. Securing edge devices (T-CNN Auto Encoder)	1. FLR improvement helpdesk tickets (TFHub, DBSCAN) 2. Asset reuse (Word2Vec) 3. Preventing revenue leakage (LSTM & BERT)







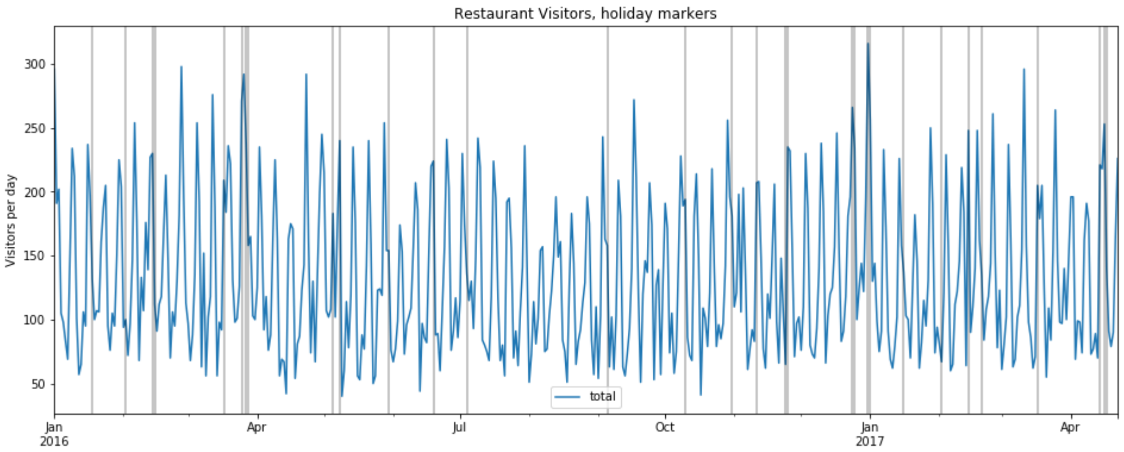
Pharma helpdesk left shift:

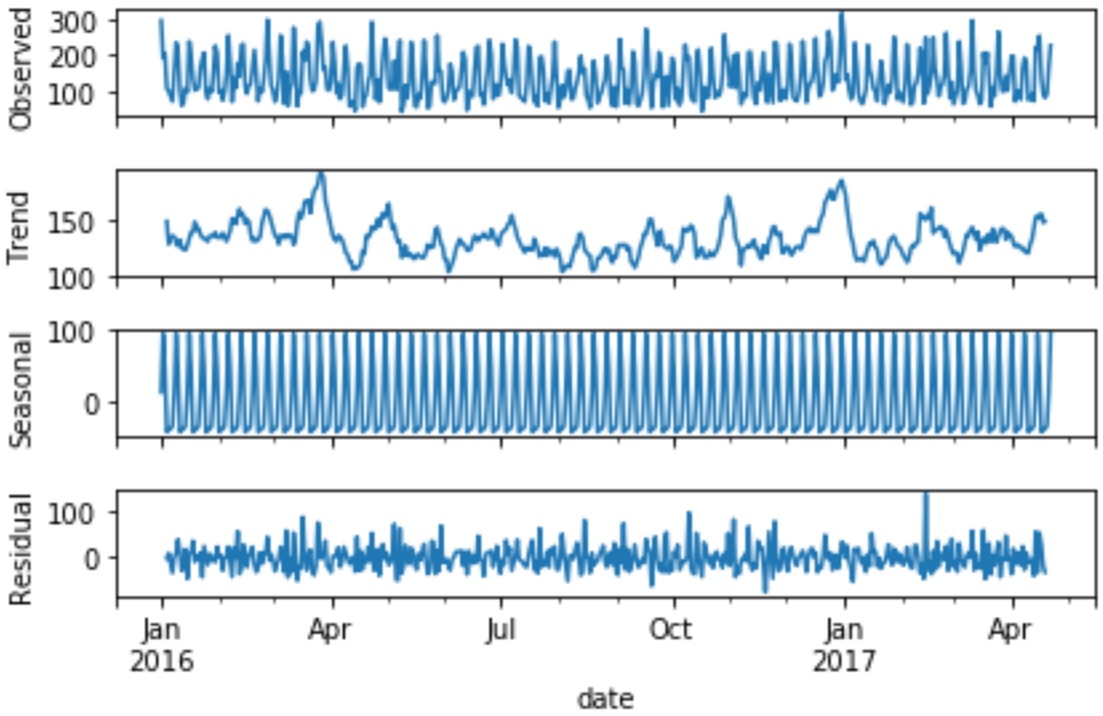
50% SME effort reduction

20 % improvement in FLR

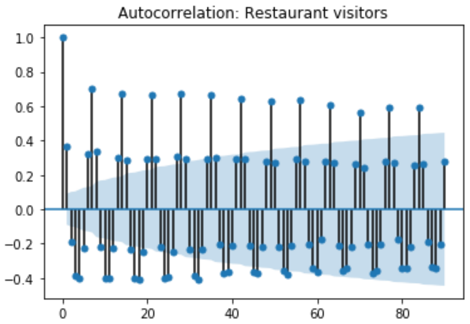
Whitepaper

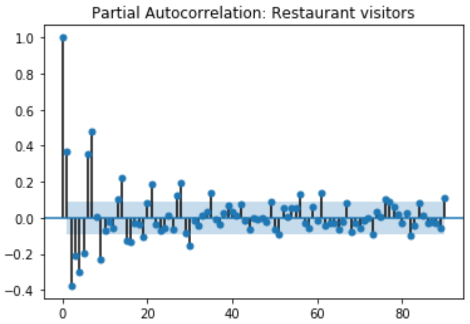
Reducing call center load, ARIMA





Augmented Dickey Fuller test for stationarity
(-5.592496972543475, 1.319377094694142e-06)

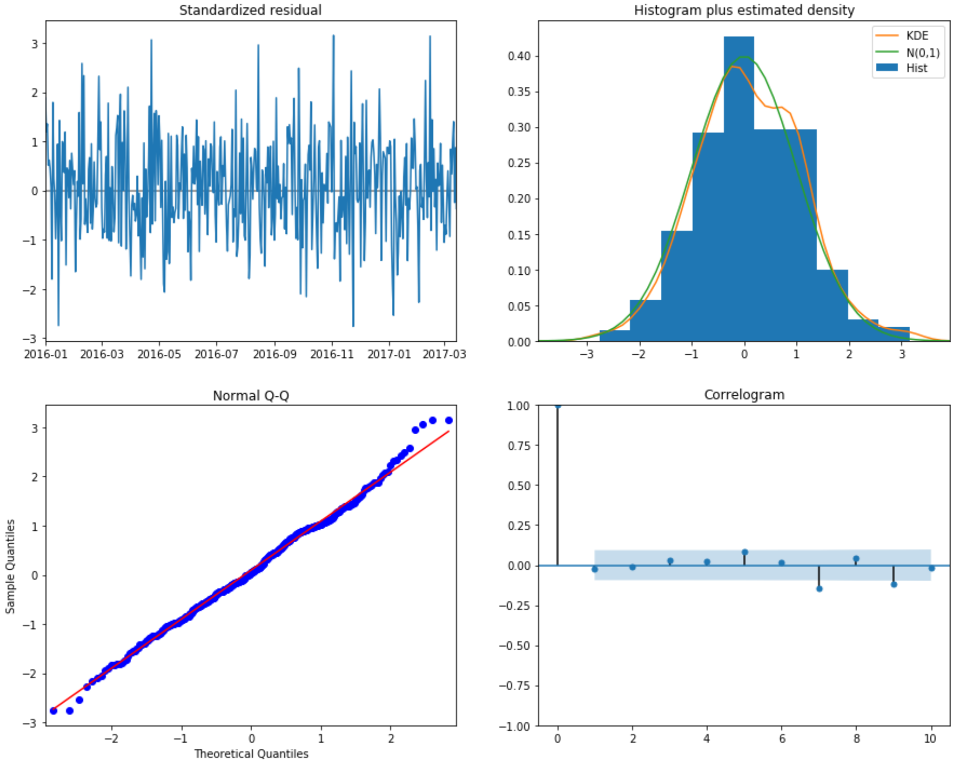


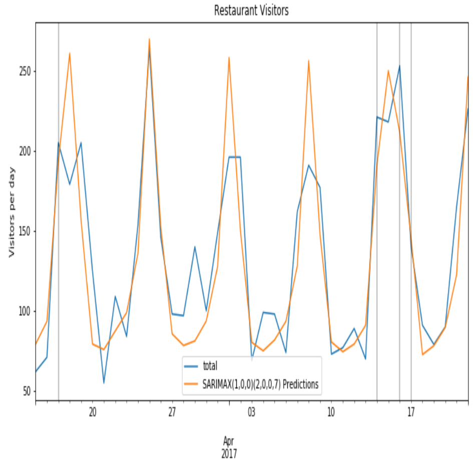


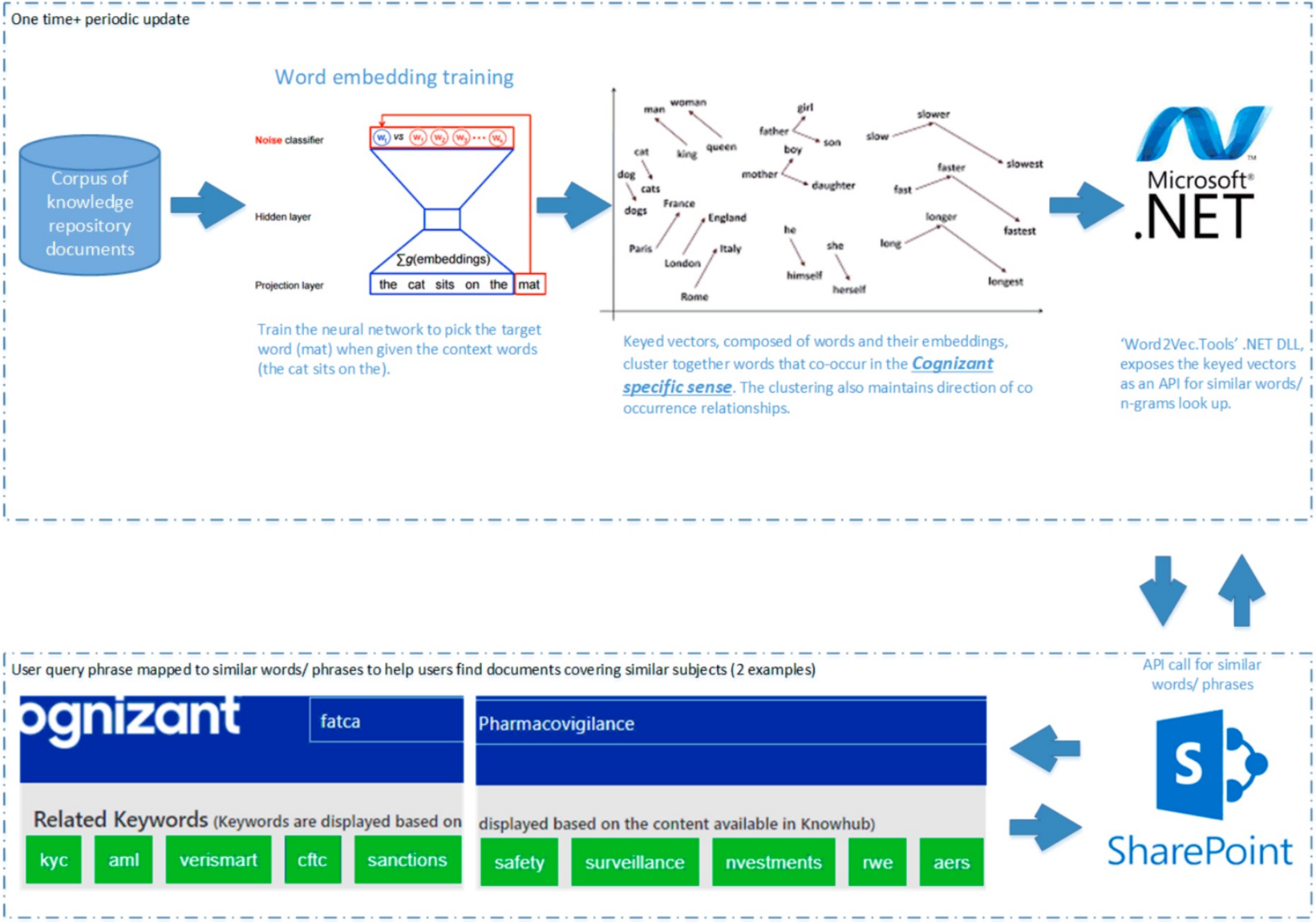
Domestic electric

- Annual seasons
- School vacations
- Zip code
- Temperature

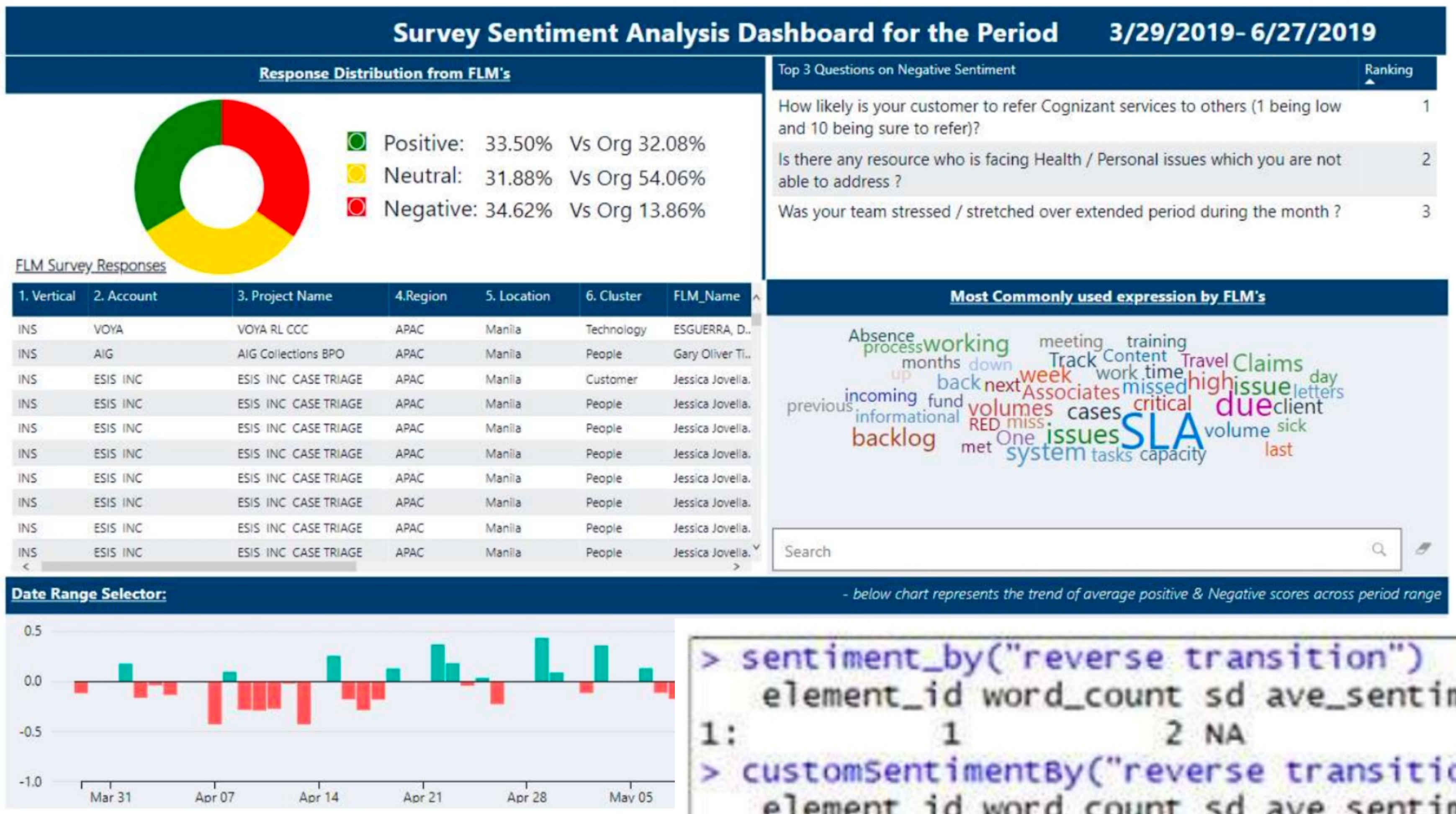
Auto_arima recco: SARIMAX(1, 0, 0)x(2, 0, 0, 7)







KnowHub: asset reuse, ~300,000 users




```
> sentiment_by("reverse transition")
   element_id word_count sd ave_sentiment
1:          1          2 NA             0
> customSentimentBy("reverse transition")
   element_id word_count sd ave_sentiment
1:          1          2 NA    -0.7071068
```

```
> sentiment_by("no developmental feedback")
   element_id word_count sd ave_sentiment
1:          1          3 NA             0
> customSentimentBy("no developmental feedback")
   element_id word_count sd ave_sentiment
1:          1          3 NA     0.5773503
```

BPO: project health measure for $17Bn p.a. projects

MACHINE HACK HACKATHONS

HOME DATA DISCUSSION SCRIPTS SUBMISSION LEADERBOARD

28. Bala Kesavan 0.96127

```
#defininig a bi-directional LSTM
inp = Input(shape=(maxlen,))
x = Embedding(max_features, embed_size, weights=[embedding_matrix])(inp)
x = Bidirectional(LSTM(50, return_sequences=True, dropout=0.05, recurrent_dropout=0.05))(x)
x = Bidirectional(LSTM(50, return_sequences=True, dropout=0.05, recurrent_dropout=0.05))(x)
x = GlobalMaxPool1D()(x)
x = Dense(50, activation="relu")(x)
x = Dropout(0.1)(x)
x = Dense(10, activation="sigmoid")(x)
model = Model(inputs=inp, outputs=x)
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

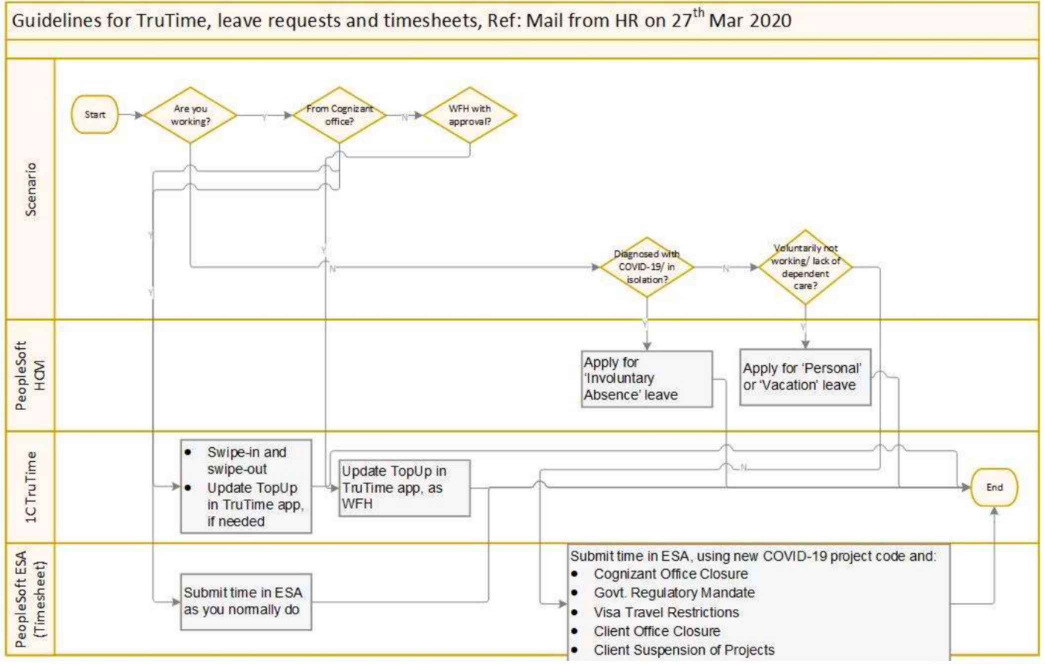
label	precision	recall	fscore	support
0	0.94	0.91	0.93	598
1	0.78	0.80	0.79	99
2	0.93	0.93	0.93	395
3	0.96	0.96	0.96	196
4	0.92	0.93	0.92	492
5	0.83	0.89	0.86	482
6	0.87	0.86	0.87	118
7	0.78	0.73	0.75	162
8	0.92	0.73	0.81	151
9	0.75	0.86	0.80	154

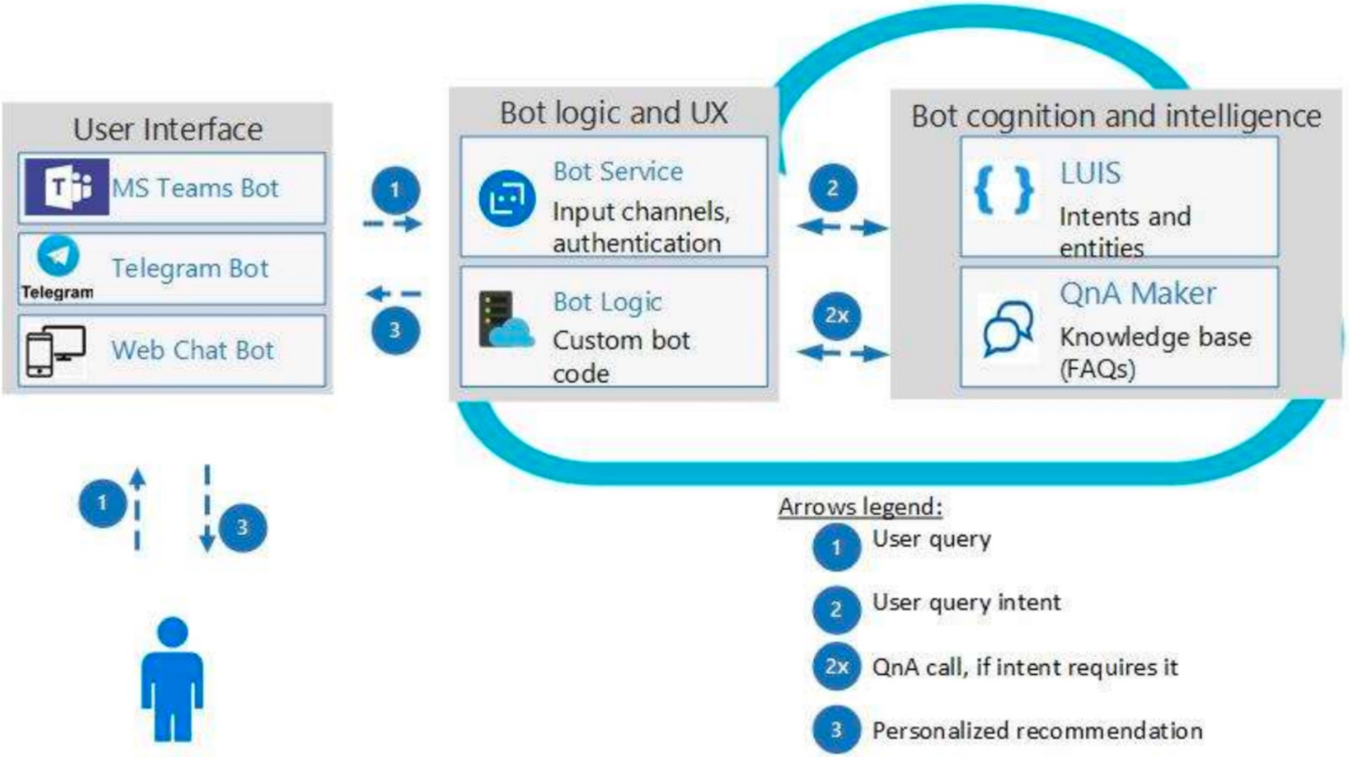
Preventing revenue leakage

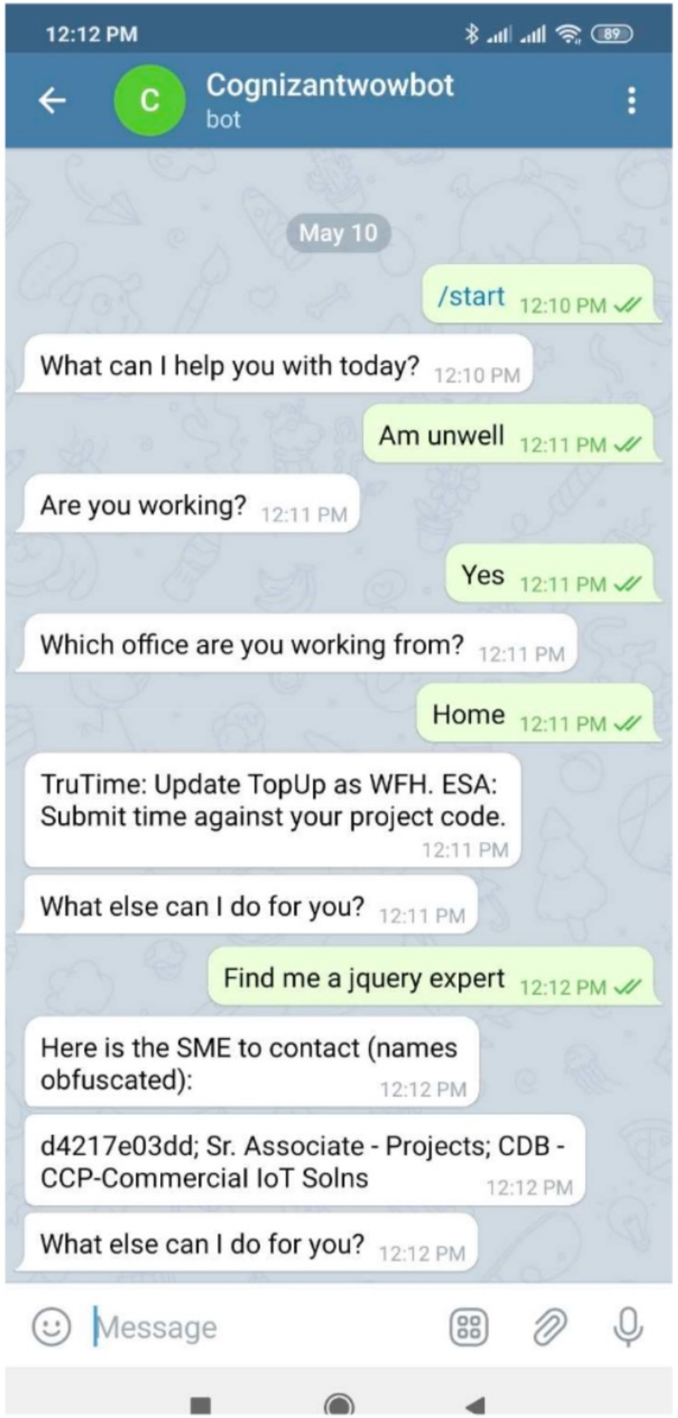
Obligations extraction from contracts

Proxy data, hackathon entry

Text classifier: BERT + TF2.0, LSTM + Keras





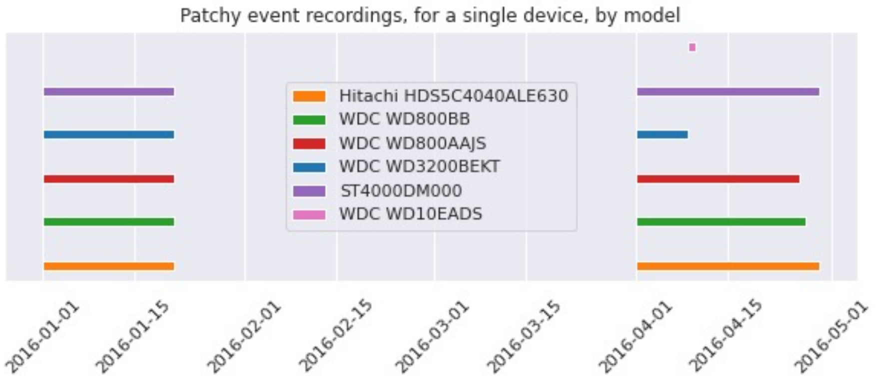


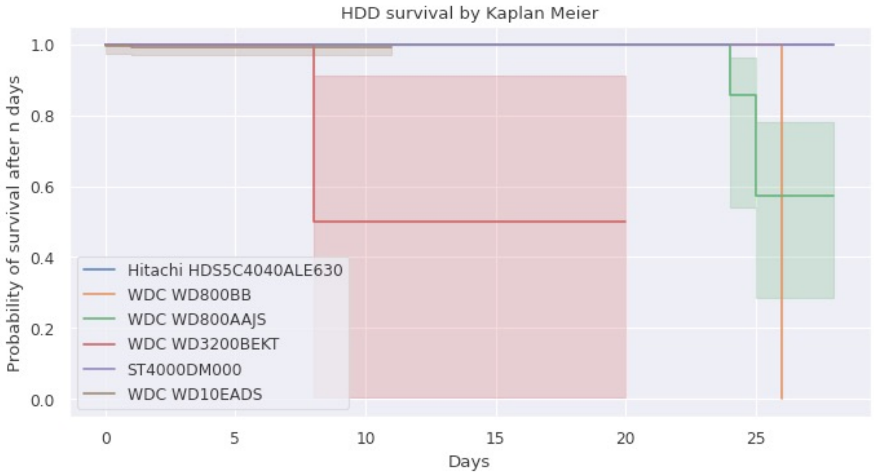
Azure helpdesk chatbot:

Employee experience

Time to roll out

Edge device failure prediction






```
These are the failures for model WDC WD3200BEKT:
          removed  observed  censored  entrance  at_risk
event_at
0.0             0         0         0         2        2
8.0             1         1         0         0        2
20.0            1         0         1         0        1
```

$$\hat{S} = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

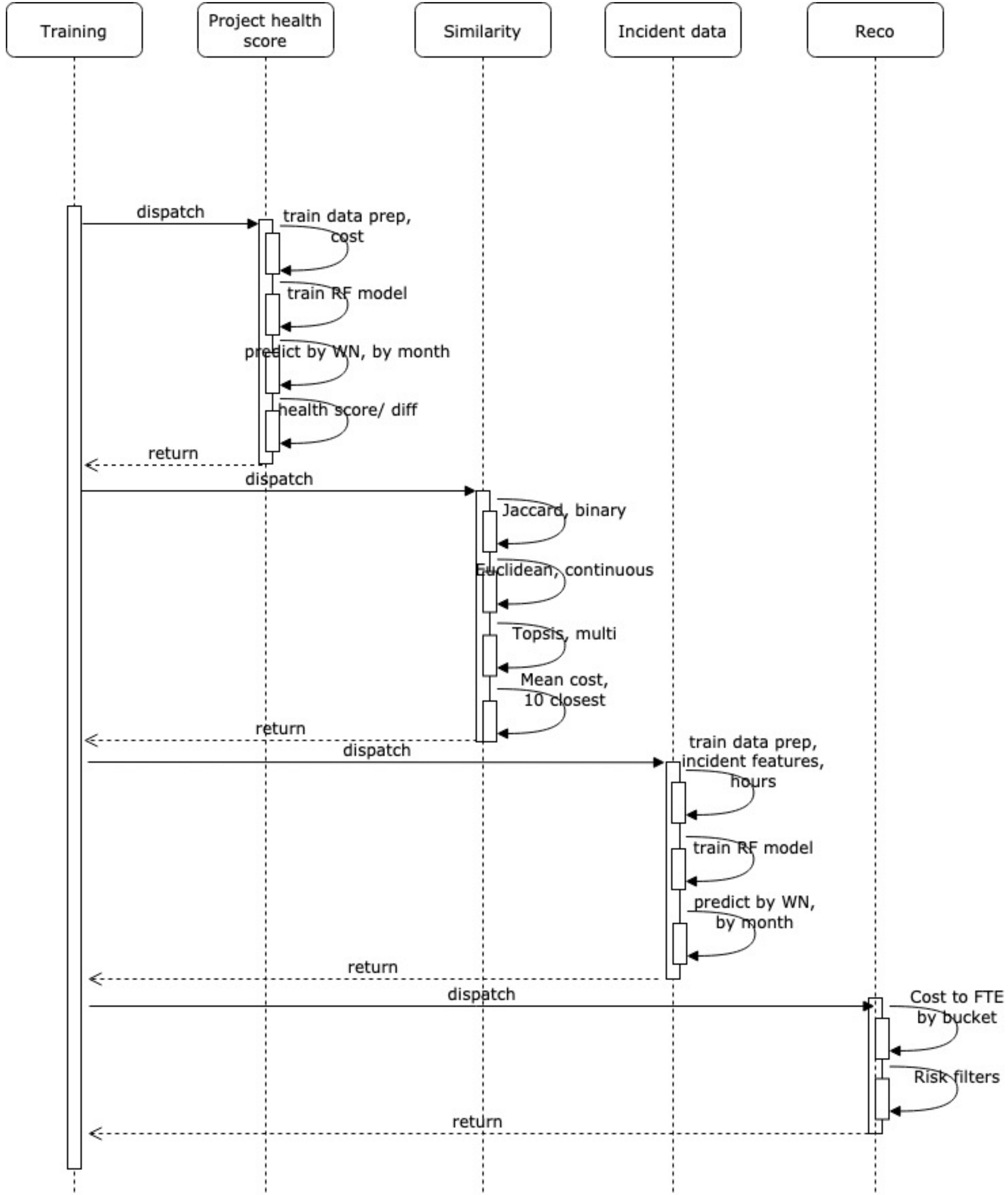
Day	Number of devices at risk	Number of devices that failed	Survival probability (product of terms)
0	2	0	(2-0)/2 = 1
8	2	1	{(2-0)/2}x{(2-1)/2} = 0.5
20	1	0	{(2-0)/2}x{(2-1)/2}x{(1-0)/1} = 0.5

t_0	-1
null_distribution	chi squared
degrees_of_freedom	1
test_name	logrank_test

		test_statistic	p	-log2(p)
Hitachi HDS5C4040ALE630	**ST4000DM000**	5.41	0.02	5.64
	WDC WD10EADS	18.72	<0.005	16.01
	WDC WD3200BEKT	751.37	<0.005	547.10
	WDC WD800AAJS	799.60	<0.005	581.94
	WDC WD800BB	262.02	<0.005	193.36
ST4000DM000	**WDC WD10EADS**	24.96	<0.005	20.70
	WDC WD3200BEKT	389.80	<0.005	285.82
	WDC WD800AAJS	621.79	<0.005	453.49
	WDC WD800BB	90.00	<0.005	68.51
WDC WD10EADS	**WDC WD3200BEKT**	13.30	<0.005	11.88
	WDC WD800AAJS	0.22	0.64	0.65
	WDC WD800BB	0.05	0.83	0.27
WDC WD3200BEKT	**WDC WD800AAJS**	14.00	<0.005	12.42
	WDC WD800BB	3.00	0.08	3.59
WDC WD800AAJS	**WDC WD800BB**	0.47	0.49	1.02

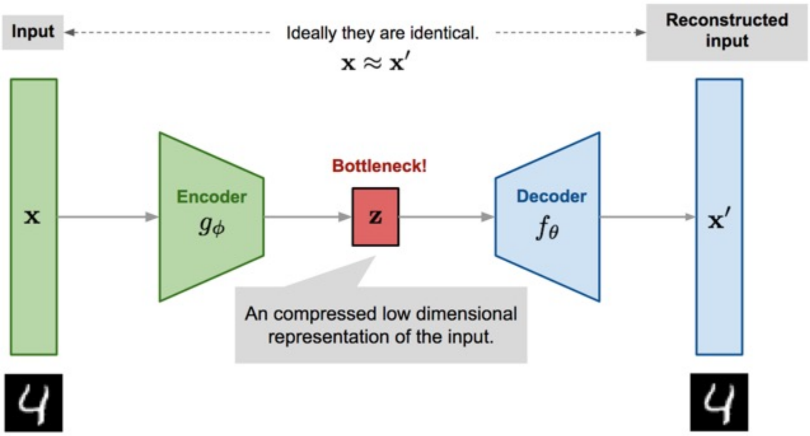
Statistical methods for automated generation of service engagement staffing plans

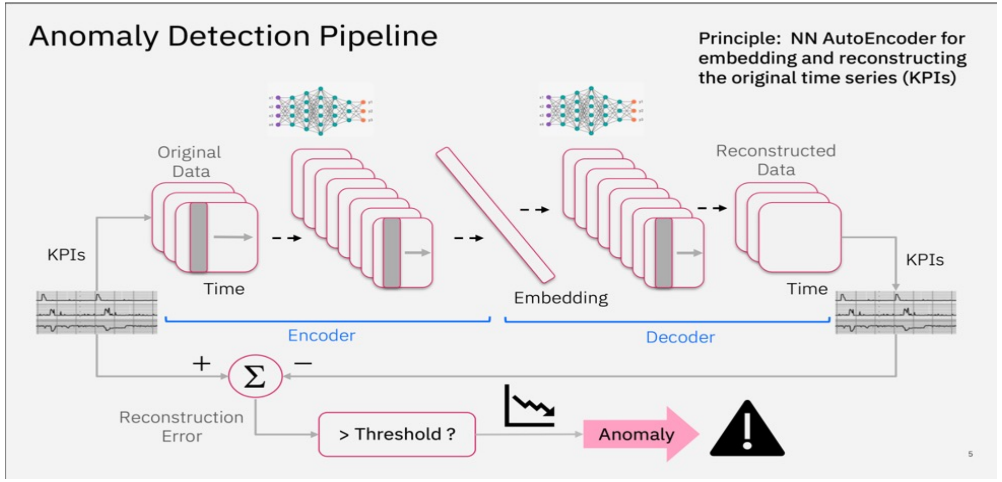
J. Hu
B. K. Ray
M. Singh

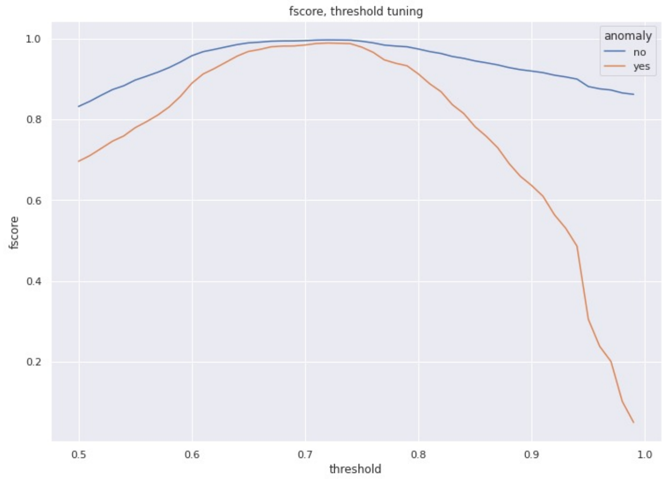


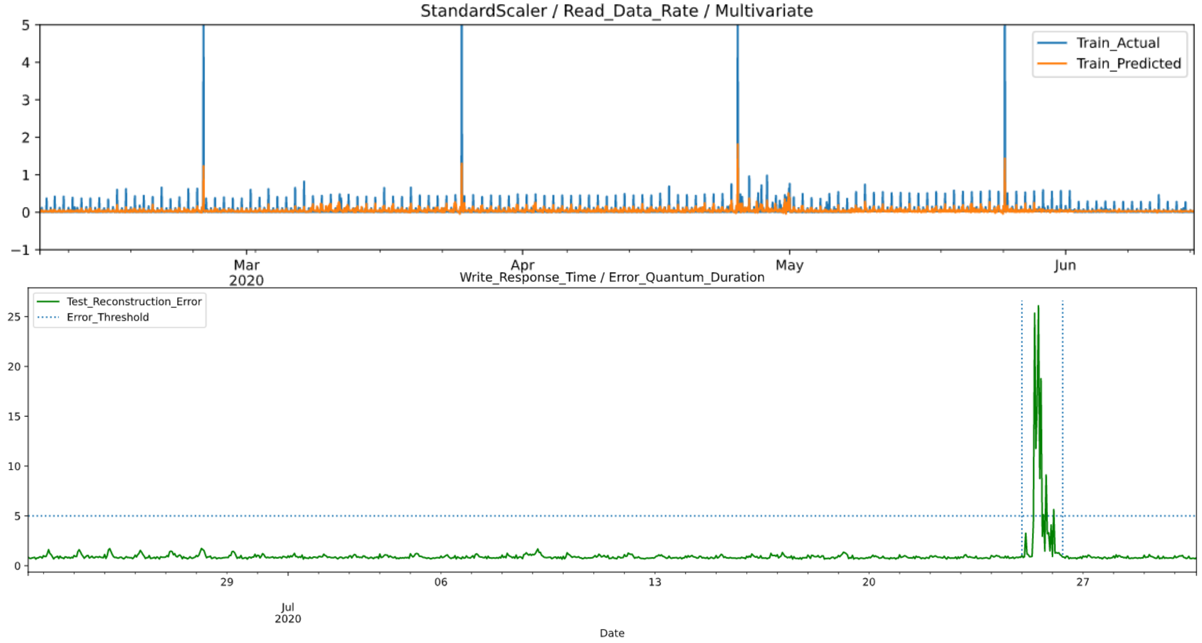
Optimizing project staffing :

Benchmark similar + well run projects









Layer (type)	Output Shape	Param #
conv1d_3 (Conv1D)	(None, 24, 64)	30784
dropout_4 (Dropout)	(None, 24, 64)	0
conv1d_4 (Conv1D)	(None, 12, 32)	12320
dropout_5 (Dropout)	(None, 12, 32)	0
conv1d_5 (Conv1D)	(None, 6, 16)	3088
conv1d_transpose_4 (Conv1DTr	(None, 12, 16)	1552
dropout_6 (Dropout)	(None, 12, 16)	0
conv1d_transpose_5 (Conv1DTr	(None, 24, 32)	3104
dropout_7 (Dropout)	(None, 24, 32)	0
conv1d_transpose_6 (Conv1DTr	(None, 48, 64)	12352
conv1d_transpose_7 (Conv1DTr	(None, 48, 80)	30800

Total params: 94,000
Trainable params: 94,000
Non-trainable params: 0

Securing edge devices (T-CNN Auto Encoder), anomaly detection