Microsoft Technology Centers

PoC Results for AIG

SOX Compliance

Microsoft

Agenda

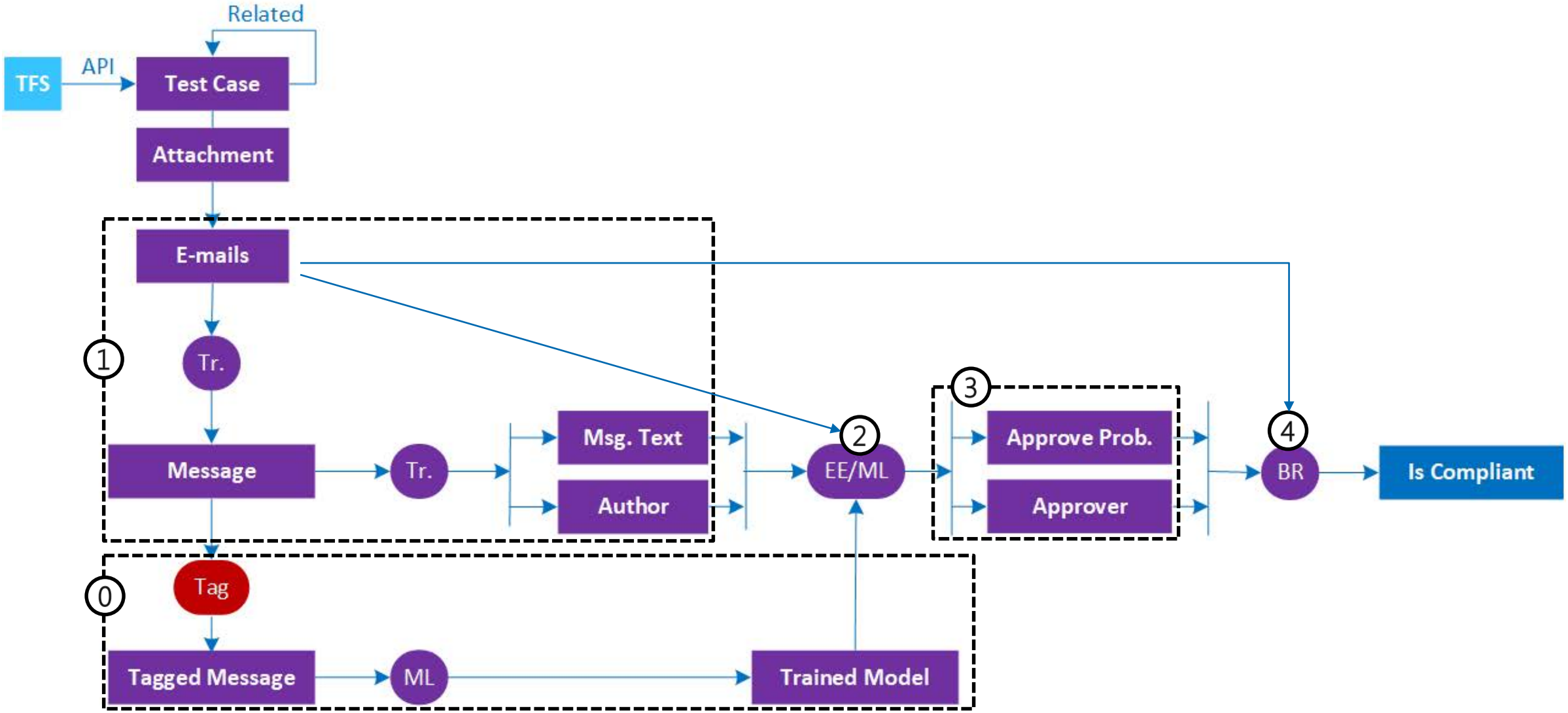
- Results of the PoC
- Review PoC status
- End-to-end workflow, created during the PoC
- Broader Workflow and Current Status
- Appendixes
- Detailed Steps of the End-to-end Workflow used in the PoC
  - 0. Train the model
  - 1. Transform e-mails to messages
  - 2. Detect Request/Approve/No-intent probability using Microsoft R
  - 3. Transform prediction to detect Approver
  - 4. Business rules
- Detect Request/Approve/No-intent probability
  - Azure ML
  - LUIS

Result of the PoC

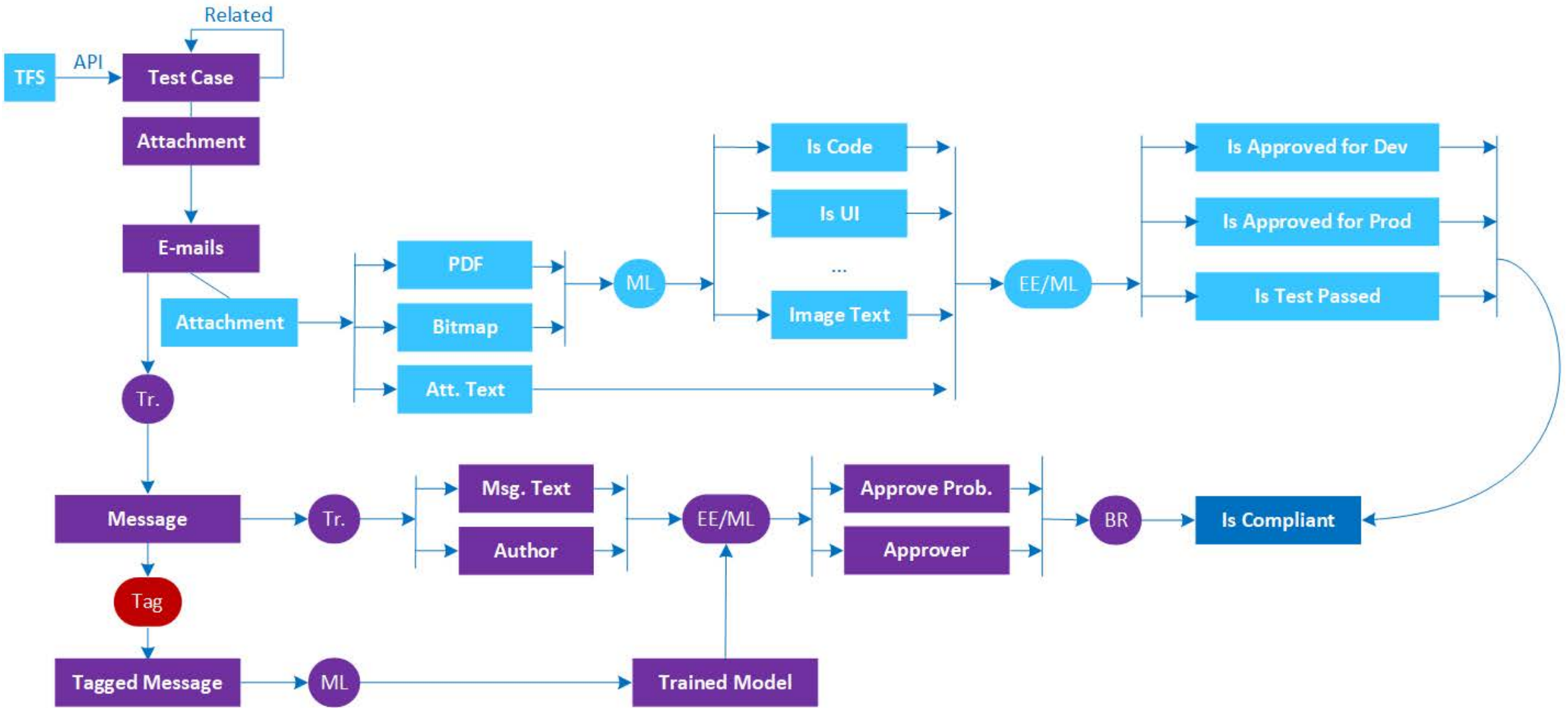
TestCas	DeployedBy	EmailApprovals
21	Prakash Muniyappa <R1-CORE\PMuniya	From: smith, Holy P Sent: Wednesday, August 16, 2017 12:28 PM To: lee, Mike B. <Mike.lee@aig.com> Subject: RE: Create new oracle data based for data science learning schema Excellent. All tables are correct. Please move to Production.
22	Challa, Ravishankar <R1-CORE\rchalla>	From: smith, Holy P Sent: Wednesday, August 16, 2017 10:39 AM To: lee, Mike B. <Mike.lee@aig.com> Subject: RE: Robotics Production Oracle Setup Approved.. Thanks. From: lee, Mike B. Sent: Wednesday, August 16, 2017 10:25 AM To: smith, Holy P <Holy.smith@aig.com> Subject: Robotics Production Oracle Setup I completed the testing on the Robotics Production Oracle Setup. This looks good fro implementation for next week. Need your approval
23	Rajendran, Pradeep <R1-CORE\PaRajenc	From: Johnson, Hoang bruce Sent: Wednesday, August 16, 2017 10:36 AM
24	Devanagi, Parashuram <R1-CORE\PDeva	From: Johnson, Hoang bruce Sent: Wednesday, August 16, 2017 11:08 AM
25	Anil Gopal Saraswathi <R1-CORE\AGopa	From: smith, Holy P Sent: Wednesday, August 16, 2017 12:44 PM
26	Devanagi, Parashuram <R1-CORE\PDeva	From: lee, Mike B.
27	Brahmaji, Gopalam <R1-CORE\gbrahma	From: smith, Holy P Sent: Wednesday, August 16, 2017 12:38 PM
28	lee, Mike B.	From: lee, Mike B. Sent: Wednesday, August 16, 2017 11:57 AM
29	Devanagi, Parashuram <R1-CORE\PDeva	From: lee, Mike B. Sent: Wednesday, August 16, 2017 11:57 AM
30	ServiceTFRMAdminPrd	
51	lee, Mike B.	From: lee, Mike B. Sent: Wednesday, August 16, 2017 11:57 AM

TestCaseNumber	DeployedBy	Approver	ApprovalProb	isCompliant	Reason
21	Prakash Muniyappa <R1-CORE\PMuniyap>	smith, Holy P	1.00	TRUE	Change management test is passed
22	Challa, Ravishankar <R1-CORE\rchalla>	smith, Holy P	1.00	TRUE	Change management test is passed
23	Rajendran, Pradeep <R1-CORE\PaRajend>	Johnson, Hoang bruce	1.00	TRUE	Change management test is passed
24	Devanagi, Parashuram <R1-CORE\PDevanag>	Johnson, Hoang bruce	1.00	TRUE	Change management test is passed
25	Anil Gopal Saraswathi <R1-CORE\AGopalsa>	smith, Holy P	0.01	FALSE	Positive intent of approval is not detected
26	Devanagi, Parashuram <R1-CORE\PDevanag>	lee, Mike B.	1.00	TRUE	Change management test is passed
27	Brahmaji, Gopalam <R1-CORE\gbrahmaj>	smith, Holy P	0.00	FALSE	Positive intent of approval is not detected
28	lee, Mike B.	lee, Mike B.	1.00	FALSE	Deployed and approved by the same person
29	Devanagi, Parashuram <R1-CORE\PDevanag>	lee, Mike B.	0.00	FALSE	Positive intent of approval is not detected
51	lee, Mike B.	lee, Mike B.	0.00	FALSE	Deployed and approved by the same person
54	Smith, Holy	smith, Holy P	1.00	TRUE	Change management test is passed
55	Potturi, Chandrasekhara <R1-CORE\CPOTTURI>	smith, Holy P	1.00	TRUE	Change management test is passed
56	Johnson, Hoang bruce	Johnson, Hoang bruce	1.00	FALSE	Deployed and approved by the same person
57	Prakash Muniyappa <R1-CORE\PMuniyap>	Johnson, Hoang bruce	1.00	TRUE	Change management test is passed
58	Yeddula, Ananda <R1-CORE\ayeddula>	smith, Holy P	0.01	FALSE	Positive intent of approval is not detected
59	Rajendran, Pradeep <R1-CORE\PaRajend>	lee, Mike B.	1.00	TRUE	Change management test is passed
60	Rajendran, Senthilkumaran <R1-CORE\srajendr>	smith, Holy P	0.00	FALSE	Positive intent of approval is not detected
81	Rajendran, Pradeep <R1-CORE\PaRajend>	smith, Holy P	1.00	TRUE	Change management test is passed
82	Bhowmik, Subhas <R1-CORE\sbhowmik>	smith, Holy P	1.00	TRUE	Change management test is passed
83	Rajendran, Pradeep <R1-CORE\PaRajend>	Johnson, Hoang bruce	0.94	TRUE	Change management test is passed
84	Prakash Muniyappa <R1-CORE\PMuniyap>	Johnson, Hoang bruce	0.08	FALSE	Positive intent of approval is not detected
86	Rajendran, Pradeep <R1-CORE\PaRajend>	smith, Holy P	1.00	TRUE	Change management test is passed
87	Rajendran, Pradeep <R1-CORE\PaRajend>	lee, Mike B.	0.99	TRUE	Change management test is passed
88	Challa, Ravishankar <R1-CORE\rchalla>	lee, Mike B.	0.87	TRUE	Change management test is passed
90	Prakash Muniyappa <R1-CORE\PMuniyap>	smith, Holy P	1.00	TRUE	Change management test is passed

End-to-end workflow, created during the PoC



Broader Workflow and Current Status



Significant opportunities
for machine learning in change management

Compliance area	Test	Industry Proven Technology
Analyze testing of the change	Have users approved the test?	
	Have users provided feedback on test	
	Is development in-line with initial request	Text mining to detect
Analyze the IT development of the change		
	Is development in-line with a testing	

Conclusion

- Based on current dataset and created models, the model built using Microsoft R Server neural network is more accurate.
- Models should be re-evaluated on a bigger real dataset to choose the best one.
- In the train dataset we need equal distribution of test cases for different classes.
- To improve the model it's needed to analyze (and potentially retag) records with false positives and false negatives.
- …..Partner…Commitments…

Appendixes

Transform e-mails to messages

```r
library("readxl")
library(dplyr)
library(tidyr)
library(xlsx)

setwd("C:\\Users\\ivank\\Microsoft\\AIG & Microsoft - Documents\\Data")
TestCases <- read_excel("test data- preliminary round-to MS-v4.xlsx")

s <- strsplit (TestCases$EmailApprovals, split = "From: ")

emails <- data.frame (TestCaseNumber = rep(TestCases$TestCaseNumber,
                                           sapply(s, length)),
                      Message = unlist(s))

messages <- (((emails %>% separate (Message, c("From", "Message"), "\r\nSent: ")
             ) %>% separate (Message, c("Sent", "Message"), "\r\nTo: ")
             ) %>% separate (Message, c("To", "Message"), "\r\nSubject: ")
             )

messages <- messages[complete.cases(messages), ]

write.xlsx(messages, "messages-v4.xlsx")
```

TestCas	RequestedBy	DeploymentType	DeployedBy	EmailApprovals	Step1	Step2	Step4
5	Dhandapani, Balaji <ut1-COutzu\	Non-Prod	BGHahmaji, Gopalam <GH1-COGHE\gb	From: smith, Holy P Sent: Wednesday, August 16, 2017 12:44 PM To: lee, Mike B. <Mike.lee@aig.com> Subject: RE: Create new oracle data based for data science learning schema Holy Smith DB Manager – IT Corporate Systems | Global - IT Tel +1 203 465 3456 | Cell +1 203 893 111 From: lee, Mike B. Sent: Wednesday, August 16, 2017 11:50 AM To: smith, Holy P <Holy.smith@aig.com> Subject: RE: Create new oracle data based for data science learning	fail	Pass	
7	smith, Holy P	Prod	uhevanagi, PaGHashuGHam <GH1-COG	From: smith, Holy P	Fail	Pass	
9	Spady, Gail L <ut1-COutzu\gaspa	Prod	uheivasigamani, GHamgesh <GH1-COG	From: lee, Mike B.	fail	Pass	
10	Aiswautya AG	Prod	GHajaGHajan SunuhaGHamuGHthy <GH1-COGHE\GHaSunuhaGH>		Fail	fail	

RowID	TestCase	From	Sent	To	Message	Intent	isApproval
39	14	lee, Mike E	Wed	Joh	Finance Asset- Import FX rates Hey give me your approval. Can't implement	Request	not Approval
41	15	smith, Hol	Wed	lee	RE: Create new oracle data based for data science	None	not Approval
42	15	lee, Mike E	Wed	sm	RE: Create new oracle data based for data science	Request	not Approval
43	15	lee, Mike E	Wed	sm	Create new oracle data based for data science	Request	not Approval
45	16	lee, Mike E	Wed	Joh	Move schema objects from abc to NBC Approved But I have never had to approve a ticket like this before- is	Approval	Approval
46	16	lee, Mike E	Wed	Joh	Move schema objects from abc to NBC Schema is ok	Request	not Approval

Detect Request/Approve/No-intent probability

Microsoft R (v.9.1)

```
#save sentiment from the pre-trained model based on "Message" as a feature
trainXdf <- RxXdfData("E:/WorkSpace/AIG_POC/train.xdf")
testXdf <- RxXdfData("E:/WorkSpace/AIG_POC/test.xdf")
rxFeaturize(data=trainData, outData = trainXdf,
            mlTransforms = list(getSentiment(vars = c(preSentiment="Message"))),
            overwrite = TRUE, randomSeed = 1
           )

rxFeaturize(data=testData, outData = testXdf,
            mlTransforms = list(getSentiment(vars = c(preSentiment="Message"))),
            overwrite = TRUE, randomSeed = 1
           )

#Define model formula and ml transformation
modelFormula <- Intent ~ MessageTrain + preSentiment
ft <- list(featurizeText(vars=c(MessageTrain="Message"), language = "English",
                         stopwordsRemover = stopwordsDefault(),
                         case = "lower", keepDiacritics = FALSE, keepPunctuations = FALSE,
                         keepNumbers = TRUE, dictionary = NULL,
                         wordFeatureExtractor = ngramCount(), charFeatureExtractor = NULL,
                         vectorNormalizer = "l2"))

columnNames <- rxGetVarNames(testData)
variables_to_remove <- c("RowID", "X")
columnNames <- columnNames[!(columnNames %in% variables_to_remove)]

#train logistic regression model
model_logistic <- rxLogisticRegression(modelFormula, trainXdf, type='multiClass', mlTransforms = ft)

#get the accuracy of logistic regression model on training and testing data
score_logistic_train <- rxPredict(model_logistic, trainXdf, extraVarsToWrite = "Intent")
score_logistic_test <- rxPredict(model_logistic, testXdf, extraVarsToWrite = columnNames)
result_logistic_train <- rxCrossTabs(~Intent:PredictedLabel, score_logistic_train, returnXtabs = TRUE)
accuracy_logistic_train <- sum(diag(result_logistic_train))/sum(result_logistic_train)
result_logistic_test <- rxCrossTabs(~Intent:PredictedLabel, score_logistic_test, returnXtabs = TRUE)
accuracy_logistic_test <- sum(diag(result_logistic_test))/sum(result_logistic_test)

#Write test results of logisitc regression model into a cvs file
write.csv(score_logistic_test, file = "E:/WorkSpace/AIG_POC/testResults_logistic.csv", row.names=FALSE)


#train neural network model
model_nn <- rxNeuralNet(modelFormula, trainXdf, type = 'multiClass', mlTransforms = ft, randomSeed = 1,
                        optimizer = adaDeltaSgd(), numIterations = 60)

#Get the accuracy of neural network model on training and testing data
score_nn_train <- rxPredict(model_nn, trainXdf, extraVarsToWrite = "Intent")
score_nn_test <- rxPredict(model_nn, testXdf, extraVarsToWrite = columnNames)
result_nn_train <- rxCrossTabs(~Intent:PredictedLabel, score_nn_train, returnXtabs = TRUE)
accuracy_nn_train <- sum(diag(result_nn_train))/sum(result_nn_train)
result_nn_test <- rxCrossTabs(~Intent:PredictedLabel, score_nn_test, returnXtabs = TRUE)
```

TestCaseN	From	Sent	To	Intent	Message	PredictedL	Score.	Score.Approval	Score.D	Score.No Intent
2	lee, Mike [	Wednesda	smith, Hol	Request	I complete	Request	0.827	0.07752908	0.0849	0.0108
3	lee, Mike [	Wednesda	Johnson, K	Request	Ready to r	Request	0.821	0.05963948	0.1014	0.0181
6	lee, Mike [	Wednesda	Johnson, K	Request	Schema is	Request	0.502	0.192264184	0.106	0.1995
15	lee, Mike [	Wednesda	smith, Hol	Request	Hate to be	Request	0.895	0.022037052	0.0648	0.0183
22	smith, Hol	Wednesda	lee, Mike [	Approval	Approved	Approval	0.042	0.911256135	0.034	0.0132
23	lee, Mike [	Wednesda	Johnson, K	Request	Ready to r	Request	0.821	0.05963948	0.1014	0.0181
27	smith, Hol	Wednesda	lee, Mike [	Disapprov	Holy COW	Disapprov	0.322	0.224204898	0.4175	0.0358
34	smith, Hol	Wednesda	lee, Mike [	Approval	Approved	Approval	0.042	0.911256135	0.034	0.0132
35	lee, Mike [	Wednesda	Johnson, K	Request	Ready to r	Request	0.821	0.05963948	0.1014	0.0181
39	lee, Mike [	Wednesda	smith, Hol	Request	FX : FX Loa	Request	0.874	0.044565924	0.0601	0.0213

Neural Networks:

```
> result_nn_train
                PredictedLabel
Intent           Request Approval Disapproval No Intent
  Request             56        0           0         0
  Approval             0       50           0         0
  Disapproval          0        0          39         0
  No Intent            3        0           0         2
> result_nn_test
                PredictedLabel
Intent           Request Approval Disapproval No Intent
  Request             12        0           0         0
  Approval             0       14           1         0
  Disapproval          0        0          11         0
  No Intent            0        0           0         0
> accuracy_nn_train
[1] 0.98
> accuracy_nn_test
[1] 0.9736842
```

Logistic Regression:

```
> result_logistic_train
                PredictedLabel
Intent           Request Approval Disapproval No Intent
  Request             54        0           2         0
  Approval             0       48           2         0
  Disapproval          0        1          38         0
  No Intent            3        2           0         0
> result_logistic_test
                PredictedLabel
Intent           Request Approval Disapproval No Intent
  Request             12        0           0         0
  Approval             0       14           1         0
  Disapproval          0        0          11         0
  No Intent            0        0           0         0
> accuracy_logistic_train
[1] 0.9333333
> accuracy_logistic_test
[1] 0.9736842
```

Microsoft Technology Centers

Transform prediction to detect Approver

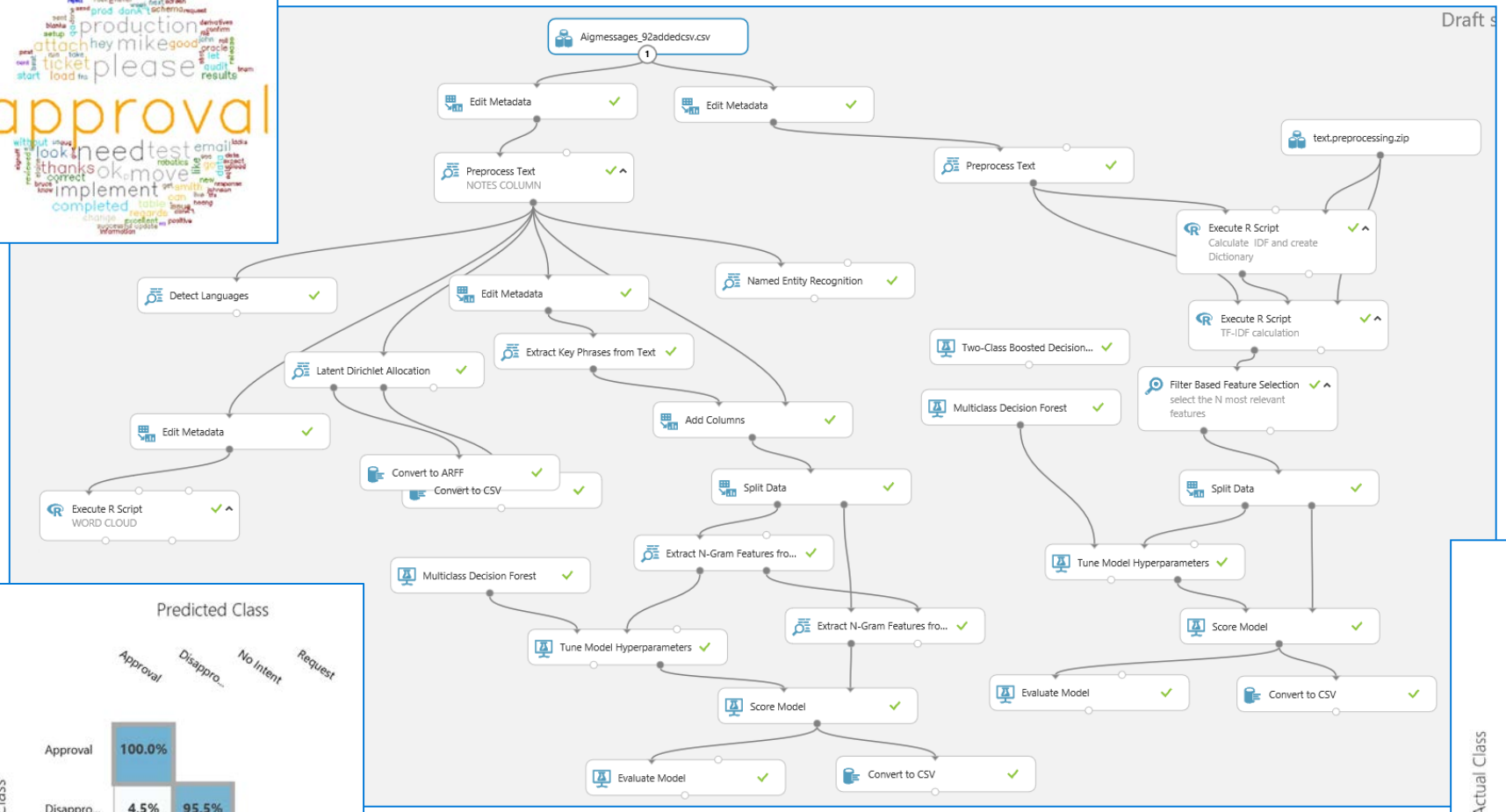
```r
# This script takes joins tickets file with emails of Approvers detected b
library("readxl")
library(dplyr)
library(tidyr)
library(xlsx)
#library(sqldf)

setwd("C:\\Users\\ivank\\Microsoft\\AIG & Microsoft - Documents\\Data")

# Read initial file
TestCases <- read_excel("test data- preliminary round-to MS-v4.xlsx")

# Read file with scored labels detected by LUIS
Scored <- read_excel("Sonia - Scored 3.xlsx")

names(Scored)[names(Scored)=="Scored.Probabilities.for.Class.Approval"] <- "ApprovalProb"
names(Scored)[names(Scored)=="Scored.Probabilities.for.Class.Disapproval"] <- "DisapprovalProb"
names(Scored)[names(Scored)=="Scored.Probabilities.for.Class.No.Intent"] <- "NoIntentProb"
names(Scored)[names(Scored)=="Scored.Probabilities.for.Class.Request"] <- "RequestProb"

# In each Ticket rank Approver by probability.
Scored <- transform(Scored, ApprovalRank =
            ave(ApprovalProb, TestCaseNumber,
                FUN = function(x) rank(-x, ties.method = "first")))

# Join TestCases and approver candidates
TestCasesApprovers <- merge (TestCases, filter (Scored, ApprovalRank == 1))

#write.xlsx(TestCasesApprovers, "tickets with approvers-v4.xlsx")
#saveWorkbook(TestCasesApprovers, "tickets with approvers-v4.xlsx")

write.csv(select(TestCasesApprovers, TestCaseNumber, From, ApprovalProb), file = "tickets with approvers-v4.csv")

```

Key Phr	RowID	TestCase	From	Se	Tc	Intent	Me:	Col	Pre	Nur	NGr	Pre	Sco	Sco	Sco	Sco	Scored Labels
fx fx loa	105	39	lee, Mik	W	sn	Request	FX : FX L		fx fx	0	[]	0	0	0	0	1	Request
produc	2	1	smith, H	W	lee	Approval	Exceller		exc	0	[]	0	1	0	0	0	Approval
ok	15	5	lee, Mik	W	sn	Request	Hate to		hat	0	[]	0	0	0	0	1	Request
release	65	24	Johnsor	W	lee	Approval	Approv		app	1	["ap	1	1	0	0	0	Approval
ok	30	11	lee, Mik	W	sn	Request	Hey test		hey	0	[]	0	0	0	0	1	Request
comple	230	91	lee, Mik	W	sn	Request	I compl		com	0	[]	0	0	0	0	1	Request
sun		141				Disapproval	Can't ap		app	1	["ap	1	0	1	0	0	Disapproval
schema	19	6	lee, Mik	W	Jo	Request	Schema		sch	0	[]	0	0	0	0	1	Request
good p	251	102	Johnsor	W	lee	Approval	Wow - t		wo	1	["ap	1	1	0	0	0	Approval
approv	39	14	lee, Mik	W	Jo	Request	Hey give		hey	0	[]	0	0	0	0	1	Request
s email	91	35	Johnsor	W	lee	Approval	Reviewe		revi	1	["ap	1	1	0	0	0	Approval
ok	154	58	lee, Mik	W	sn	Request	Hey test		hey	0	[]	0	0	0	0	1	Request

TestCaseNumber	From	ApprovalProb
1	smith, Holy P	0.82
2	smith, Holy P	0.86
4	Johnson, Hoang bruce	1.00
5	smith, Holy P	0.33
6	lee, Mike B.	0.07
8	lee, Mike B.	0.92
9	lee, Mike B.	0.65

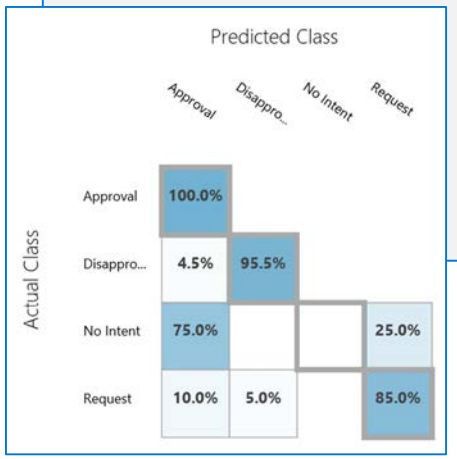
Detect Request/Approve/No-intent probability

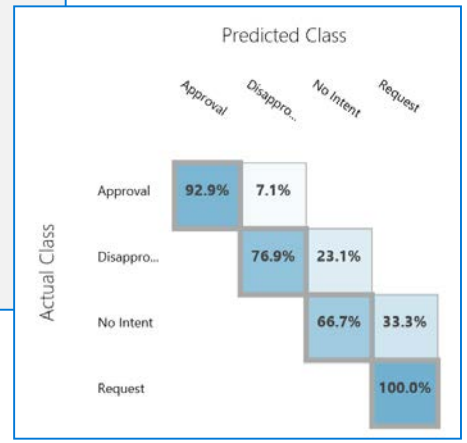
Azure ML



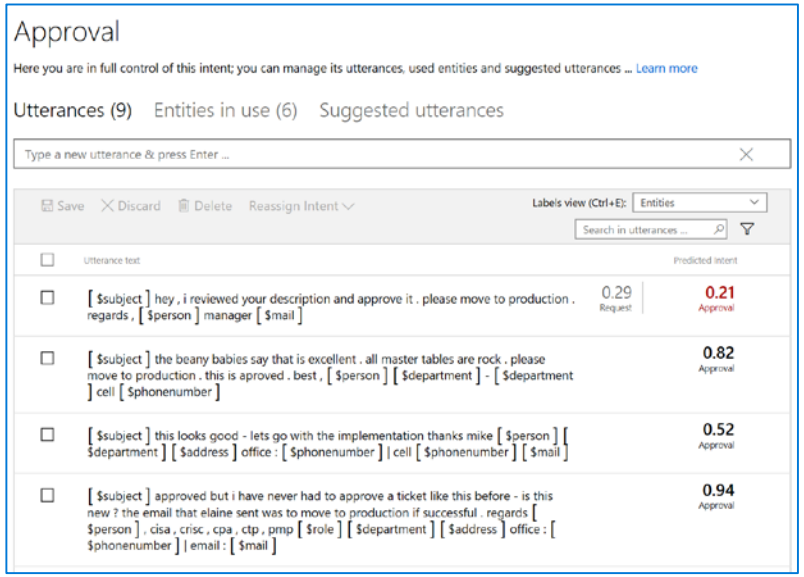
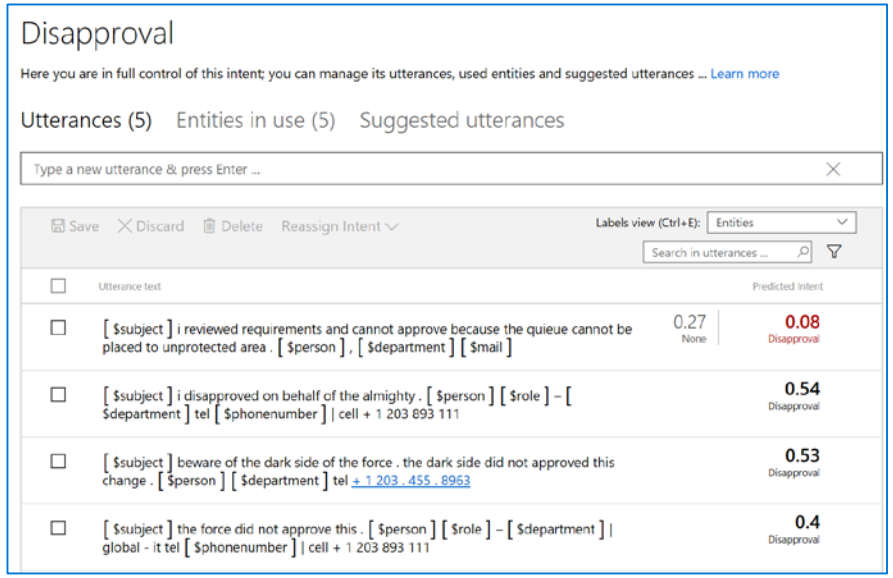
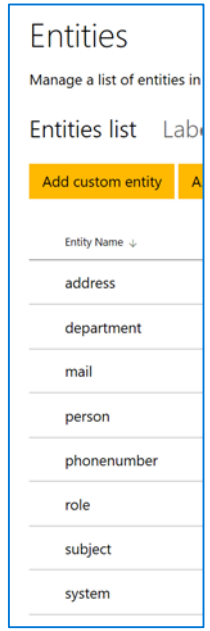
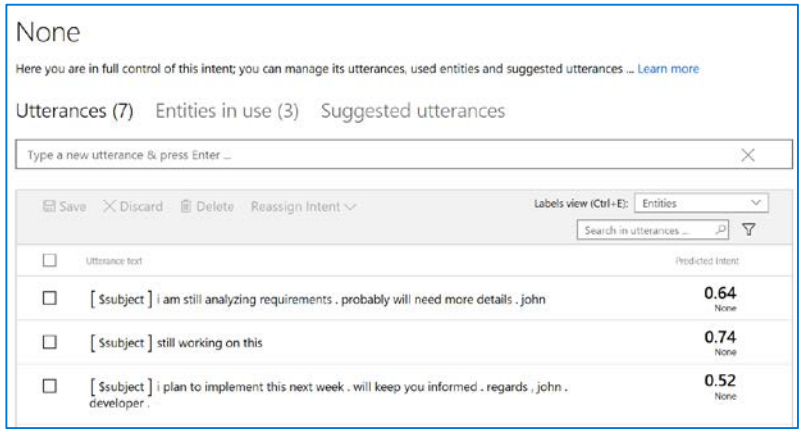
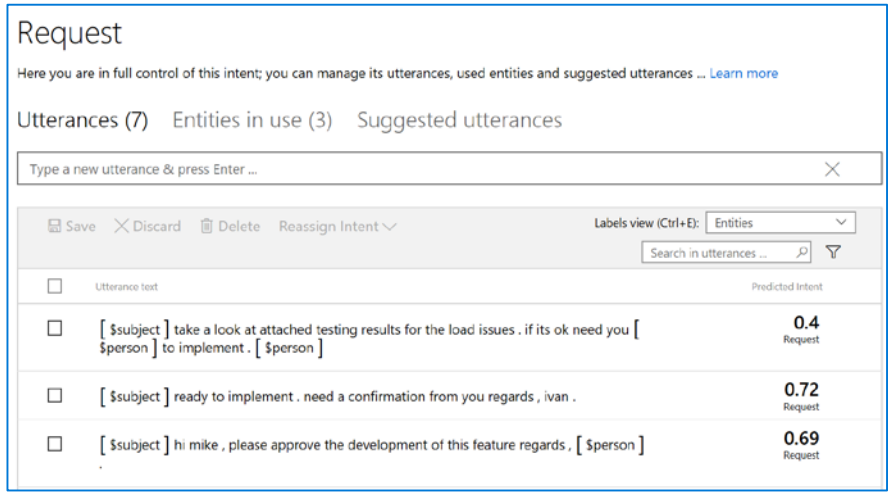
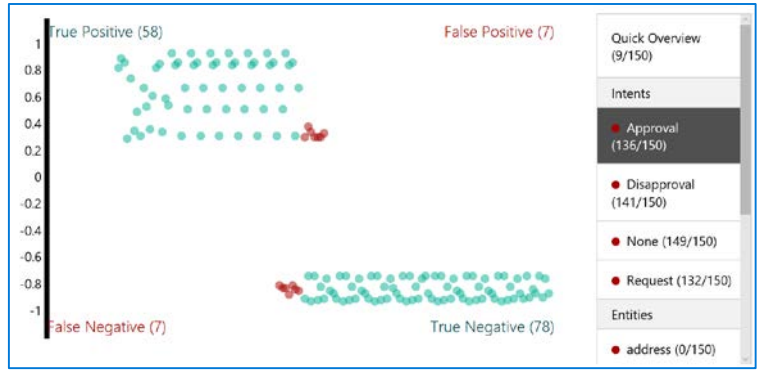
RowID	TestCase	From	Sent	To	Message	Intent	isApproval
39	14	lee, Mike E	Wed	Joh	Finance Asset- Import FX rates Hey give me your approval. Can't implement	Request	not Approval
41	15	smith, Hol	Wed	lee	RE: Create new oracle data based for data science	None	not Approval
42	15	lee, Mike E	Wed	sm	RE: Create new oracle data based for data science	Request	not Approval
43	15	lee, Mike E	Wed	sm	Create new oracle data based for data science	Request	not Approval
45	16	lee, Mike E	Wed	Joh	Move schema objects from abc to NBC Approved But I have never had to approve a ticket like this before- is	Approval	Approval
46	16	lee, Mike E	Wed	Joh	Move schema objects from abc to NBC Schema is ok	Request	not Approval

Key Phr	RowID	TestCase	From	Se	To	Intent	Me	Col	Pre	Nur	NGr	Pre	Sco	Sco	Sco	Sco	Scored Labels
fx fx loa	105	39	lee, Mik	W	sn	Request	FX : FX L	fx f	0	[]	0	0	0	0	0	1	Request
produc	2	1	smith, H	W	lee	Approval	Excellen	exc	0	[]	0	1	0	0	0	0	Approval
ok	15	5	lee, Mik	W	sn	Request	Hate to	hat	0	[]	0	0	0	0	0	1	Request
release	65	24	Johnsor	W	lee	Approval	Approv	app	1	["ap	1	1	0	0	0	0	Approval
ok	30	11	lee, Mik	W	sn	Request	Hey test	hey	0	[]	0	0	0	0	0	1	Request
comple	230	91	lee, Mik	W	sn	Request	I compl	com	0	[]	0	0	0	0	0	1	Request
sun		141				Disapproval	Can't ap	app	1	["ap	1	0	1	0	0	0	Disapproval
schema	19	6	lee, Mik	W	Jo	Request	Schema	sch	0	[]	0	0	0	0	0	1	Request
good p	251	102	Johnsor	W	lee	Approval	Wow - t	wo	1	["ap	1	1	0	0	0	0	Approval
approv	39	14	lee, Mik	W	Jo	Request	Hey give	hey	0	[]	0	0	0	0	0	1	Request
s email	91	35	Johnsor	W	lee	Approval	Reviewe	revi	1	["ap	1	1	0	0	0	0	Approval
ok	154	58	lee, Mik	W	sn	Request	Hey test	hey	0	[]	0	0	0	0	0	1	Request





Microsoft Technology Centers

Detect Request/Approve/No-intent probability

LUIS