Hotel Staff Feedback

Project 3, Group 6, DSI13

Elton, Jason and Joey

Context and Problem Statement

We are the American Hotel and Lodging Association.

- In recent years, f**rontdesk staff at hotels have been leaving** due to increased workload.

- **Vicious cycle** - as frontdesk staff decrease, workload increases for remaining staff, who then leave etc.

- **Affected customer experience** at front desks.

Context and Problem Statement

We currently have access to annual feedback from hotel staff (including our servers in the ballrooms and restaurants, staff at the frontdesk etc.)

- We want to zoom in to problems from frontdesk staff only, so that we can quickly stop the vicious cycle of such staff leaving, and service standards declining.

- Production Model > 90% Accuracy!

Data Cleaning

- Removed duplicates
- Removed moderator message
- Advertisements were not found in the scraped data
- Treatment of posts with text in the title, but no text in the body - joined body (“selftext”) and title of all posts to create a new column

Preprocessing

We created a function to preprocess the text. The function helped to:

- remove any ‘http’
- remove non-letters and split the text into just words so that they can be processed
- convert all letters to lower case
- get rid of word variations of same root meaning - lemmatization
- remove stop words from a pre-determined and customized set
- re-join words back into text for EDA

EDA


room
time
just
guest
hotel
i'm
like
know
night
come
people
don't
said
tell
say
going
thing
check
server
manager
shift
table
told
want
got
desk
work
make
guy
ask
day
food
minute
right
think
it
way
phone
restaurant
order
went
job
asked
card
didn't
really
i've
hour
point
came
sure
started
floor
reservation
good
need
look
door
week

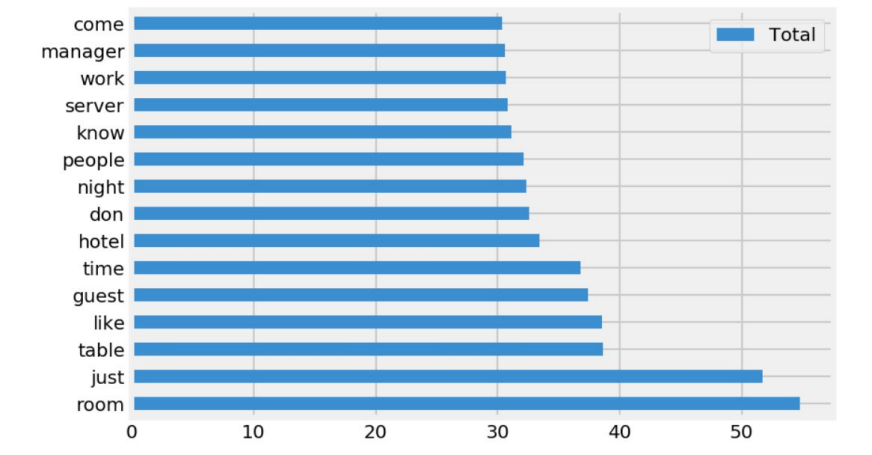

just
table
like
server
time
restaurant
manager
people
come
tip
don
know
drink
ve
ll
got
food
night
going
really
told
day
said
didn
ask
work
want
guy
bar
need
tell
ha
thing
order
say
job
customer
shift
minute
make
lady
did
getting
check
kitchen
today
left
friend
good
right
asked
new
took
sorry
week
wait
went
menu
year
wanted
coworker
person
little
started
working


guest
just
room
said
say
come
know
night
tell
told
people
time
like
hotel
check
desk
work
guy
ha
going
did
day
manager
thing
ve
got
good
need
didn
want
phone
card
place
look
let
getting
little
way
right
minute
don
stay
walk
ask
called
think
ll
leave
reallyplace
reservation
went
job
man
start
trying
came
sure
property
people

Modelling Process

- Naive Bayes and Logistic Regression.
- For each model,
  - CountVectorizer (cvec) and TfidfVectorizer (tvec) to vectorize texts into word tokens (compatible with modelling)

Eg.: Word weightage for X_train after tvec treatment:

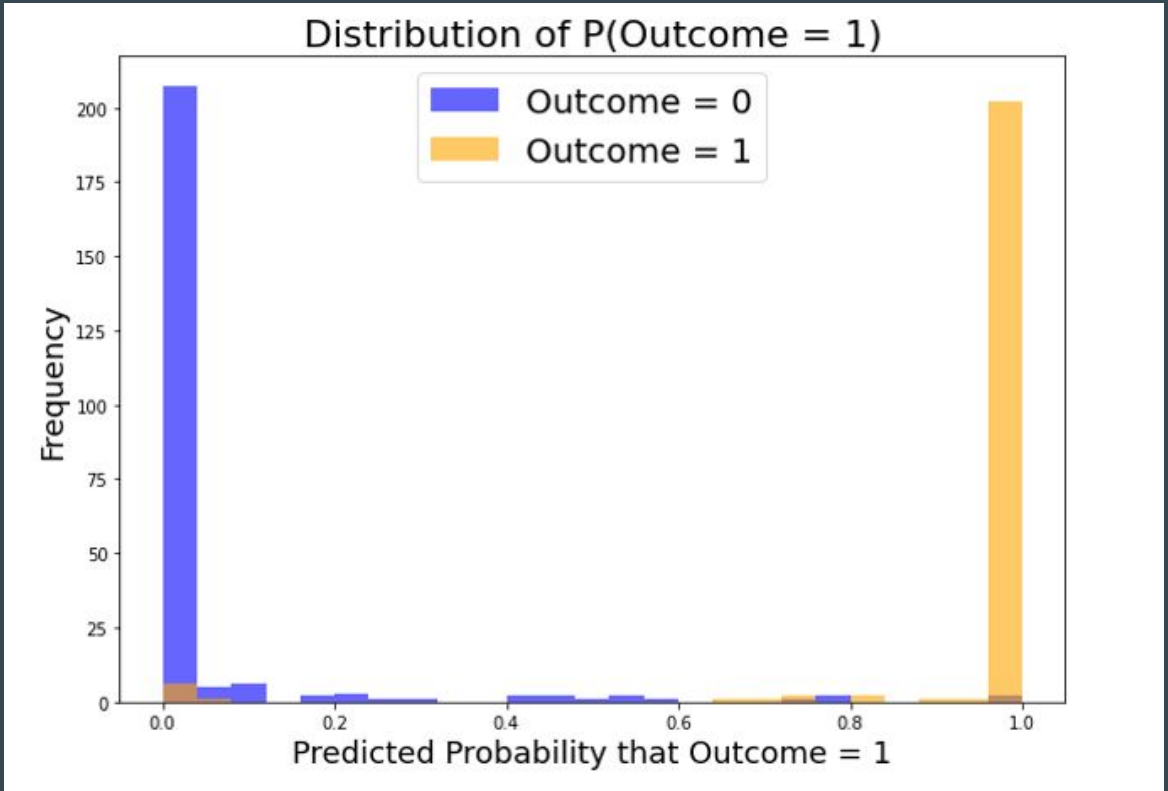


Model of Choice - Comparison of Model Results

Multinomial Naive Bayes with Count Vectorizer

Model	Train Accuracy	Test Accuracy	Sensitivity	Specificity	Precision	F1-score	ROC AUC score
CVec_LogReg	0.997	0.94	0.94	0.935	0.94	0.938	0.99
TVec_LogReg	0.986	0.965	0.94	0.987	0.986	0.962	0.994
CVec_MNB	0.966	0.967	0.968	0.966	0.963	0.966	0.986
TVec_MNB	0.965	0.965	0.959	0.971	0.967	0.963	0.99

Model of Choice - Prediction Probabilities Distribution



Model Interpretation - Visualization

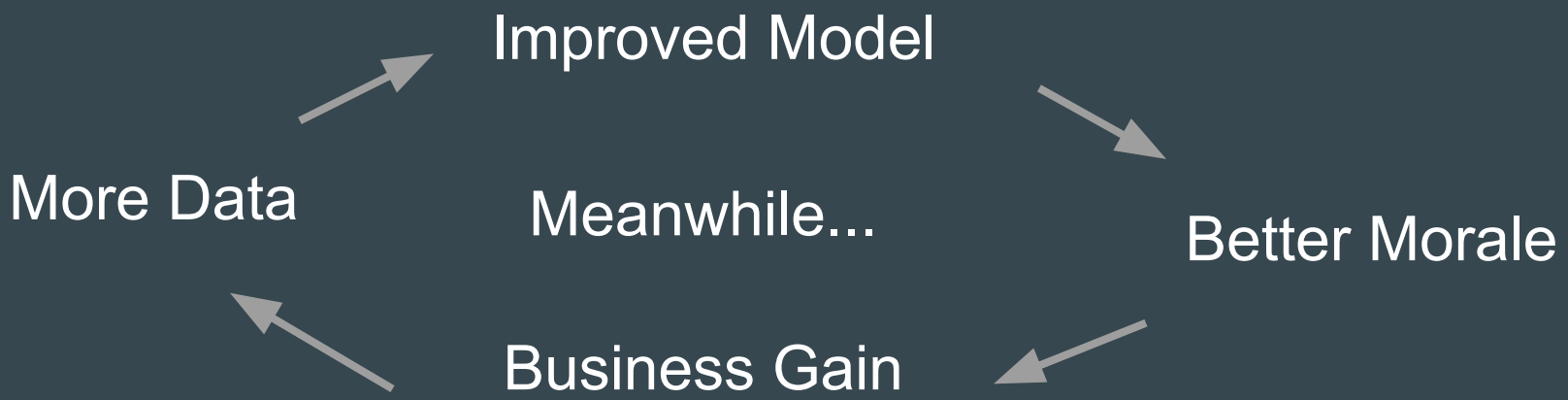
Pre-Model Wordcloud

Production Model Wordcloud

Conclusion and Recommendations

CVec_LogReg model (Accuracy 0.965) >> Baseline Model (Accuracy 0.522)

CVec_LogReg model (Accuracy 0.965) > Problem Statement (Accuracy 0.9)



Conclusion and Recommendations

CVec_LogReg model (Accuracy 0.965) >> Baseline Model (Accuracy 0.522)

CVec_LogReg model (Accuracy 0.965) > Problem Statement (Accuracy 0.9)