Exploring the Benefits of AI-Powered Insights in Business (NLP)

Ai tools.

Group Members

- Anthony Ngatia
- Elsie Nduta
- Joy kipkemboi
- Jessyca Aperi
- Rotich Naomi

Business understanding

Limited sensitivity in NLP emotion analysis impedes empathetic interactions. Leveraging the GoEmotions dataset and neural networks/SVM models for improved emotion classification and user experiences.

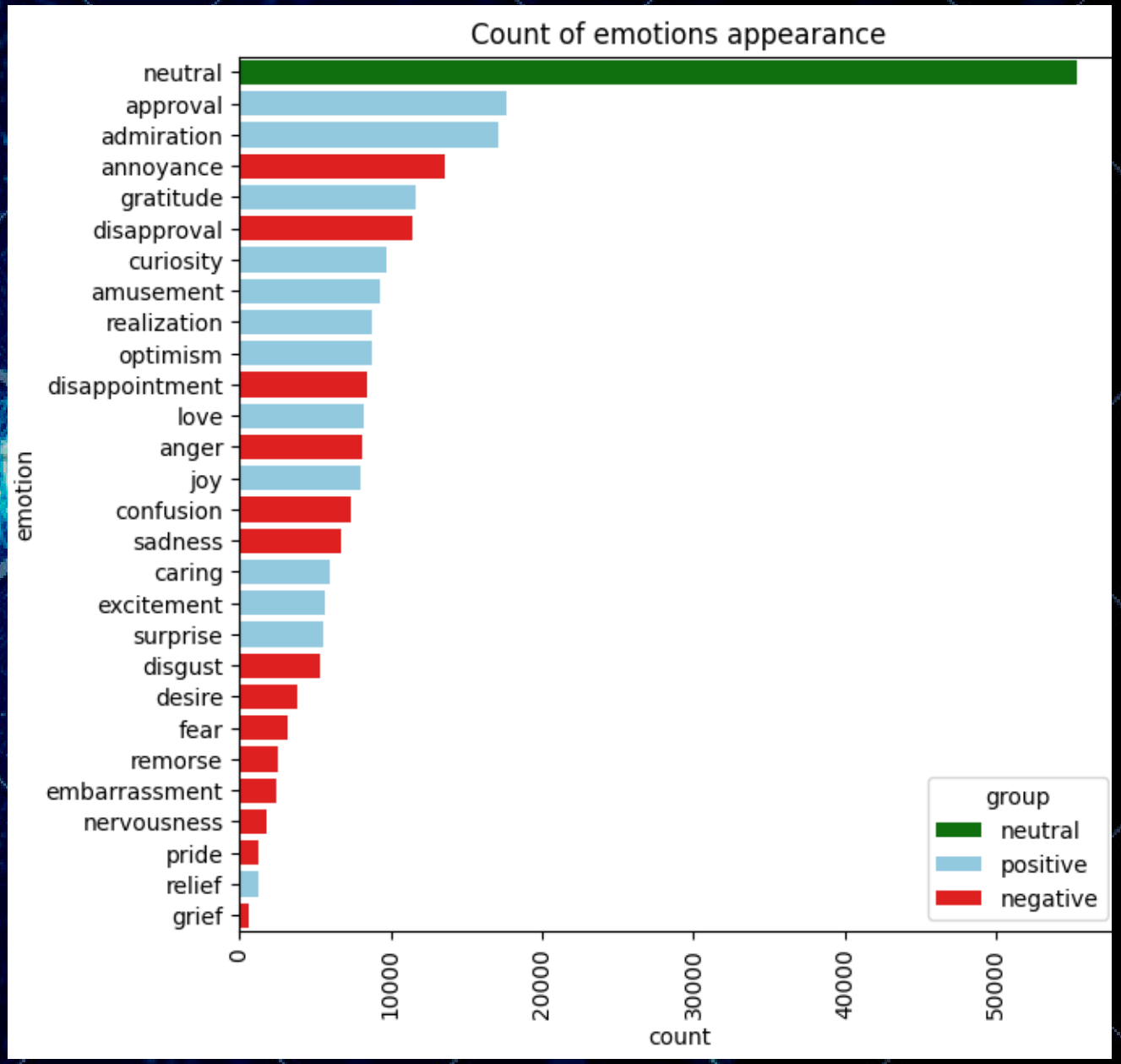
Main Objective

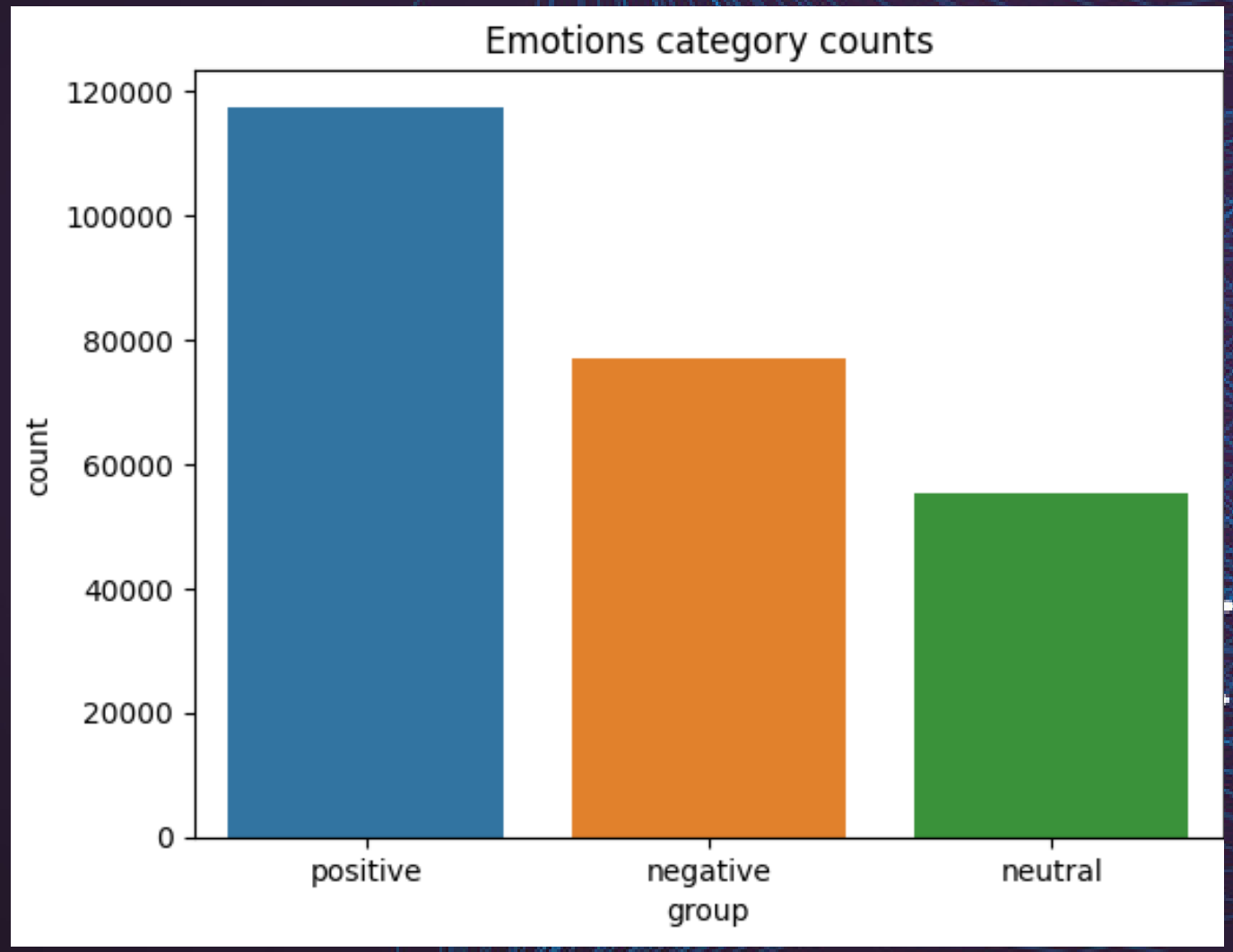
Expand emotion classification datasets by training models to analyze text tonality using the Google AI GoEmotions dataset.

Specific objectives

- Improve customer support by recognizing and addressing user emotions in textual communication.
- Develop a model to classify and accurately predict sentiments into different sentiment categories

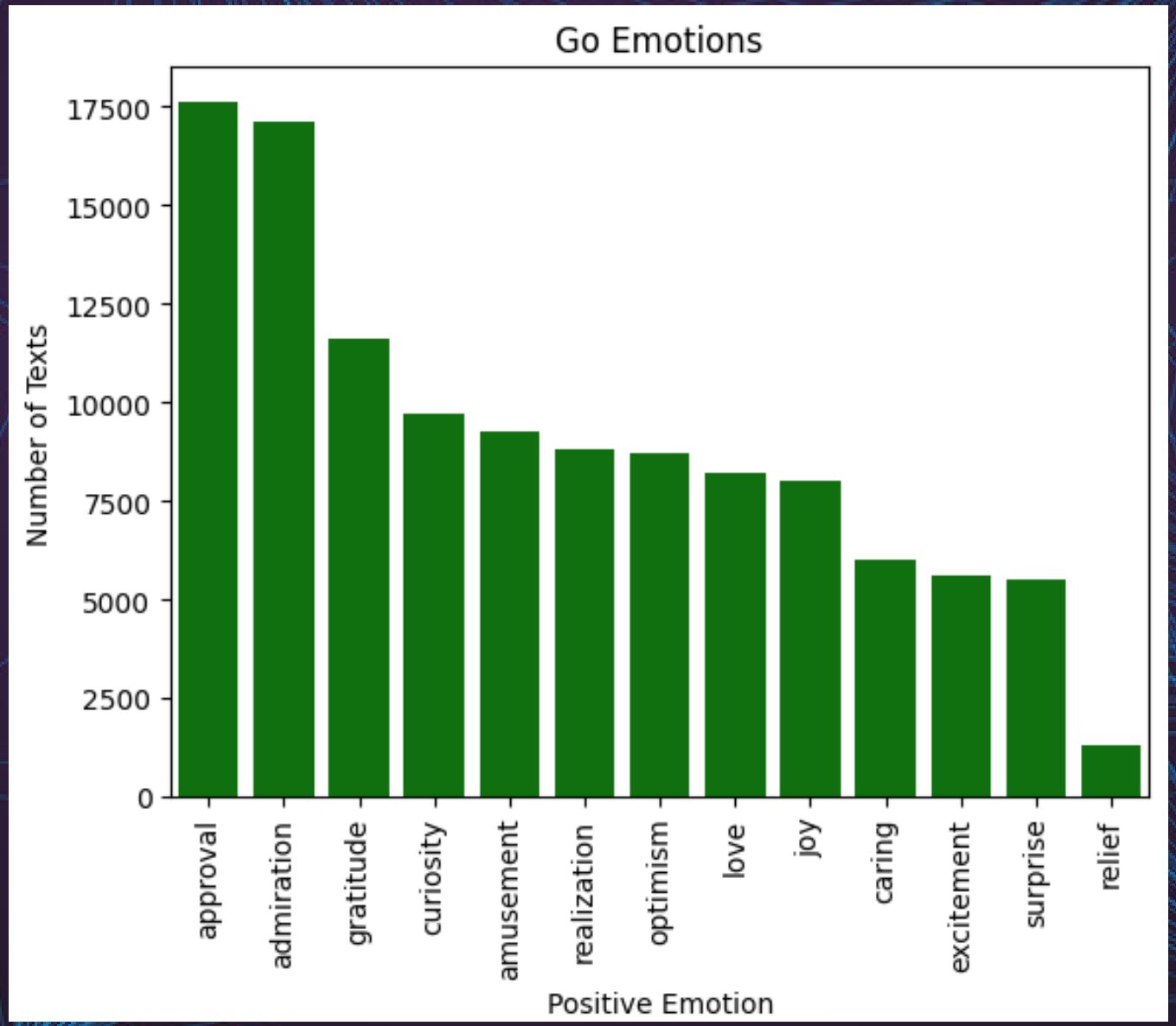
Exploratory Data Analysis



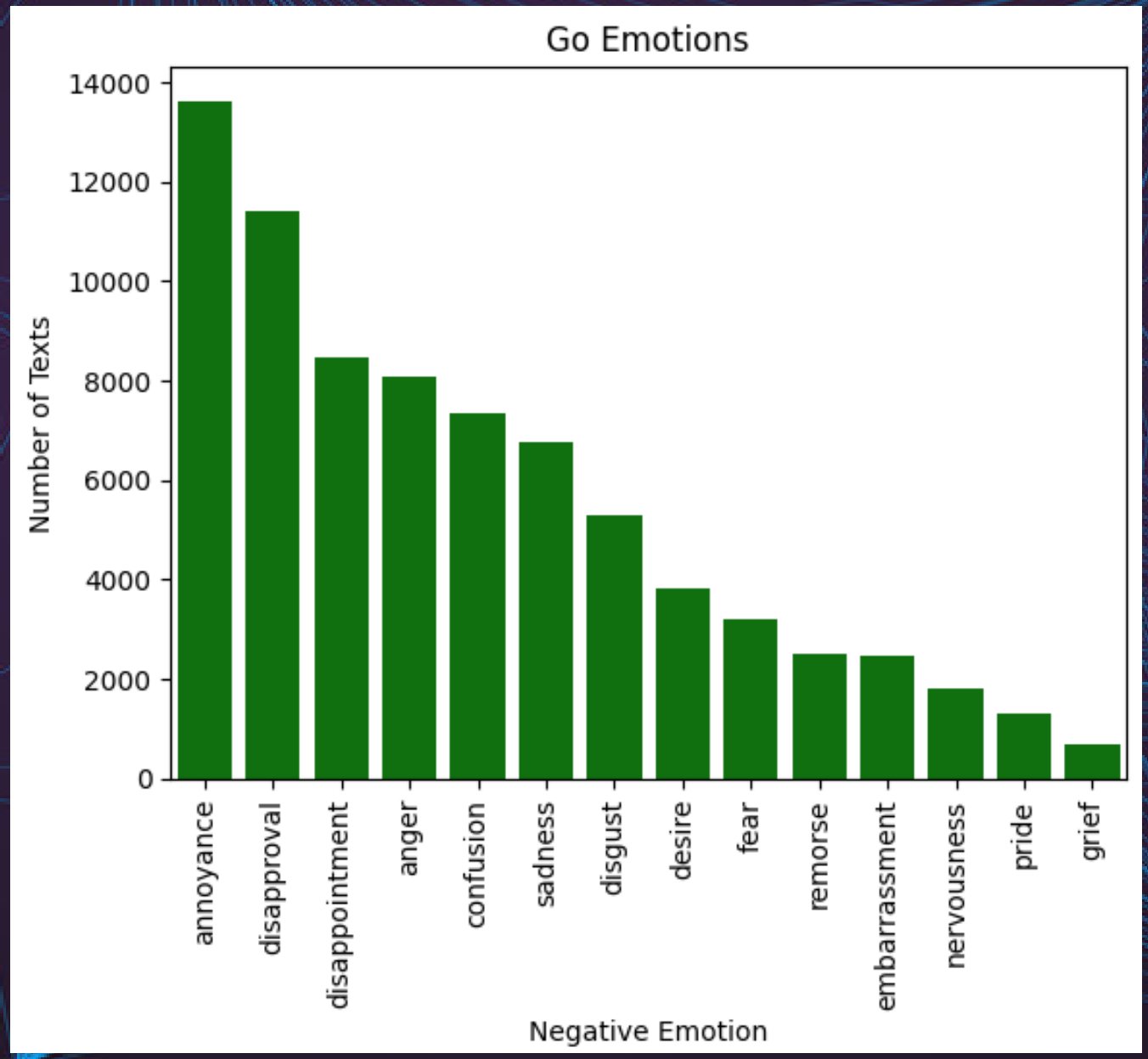


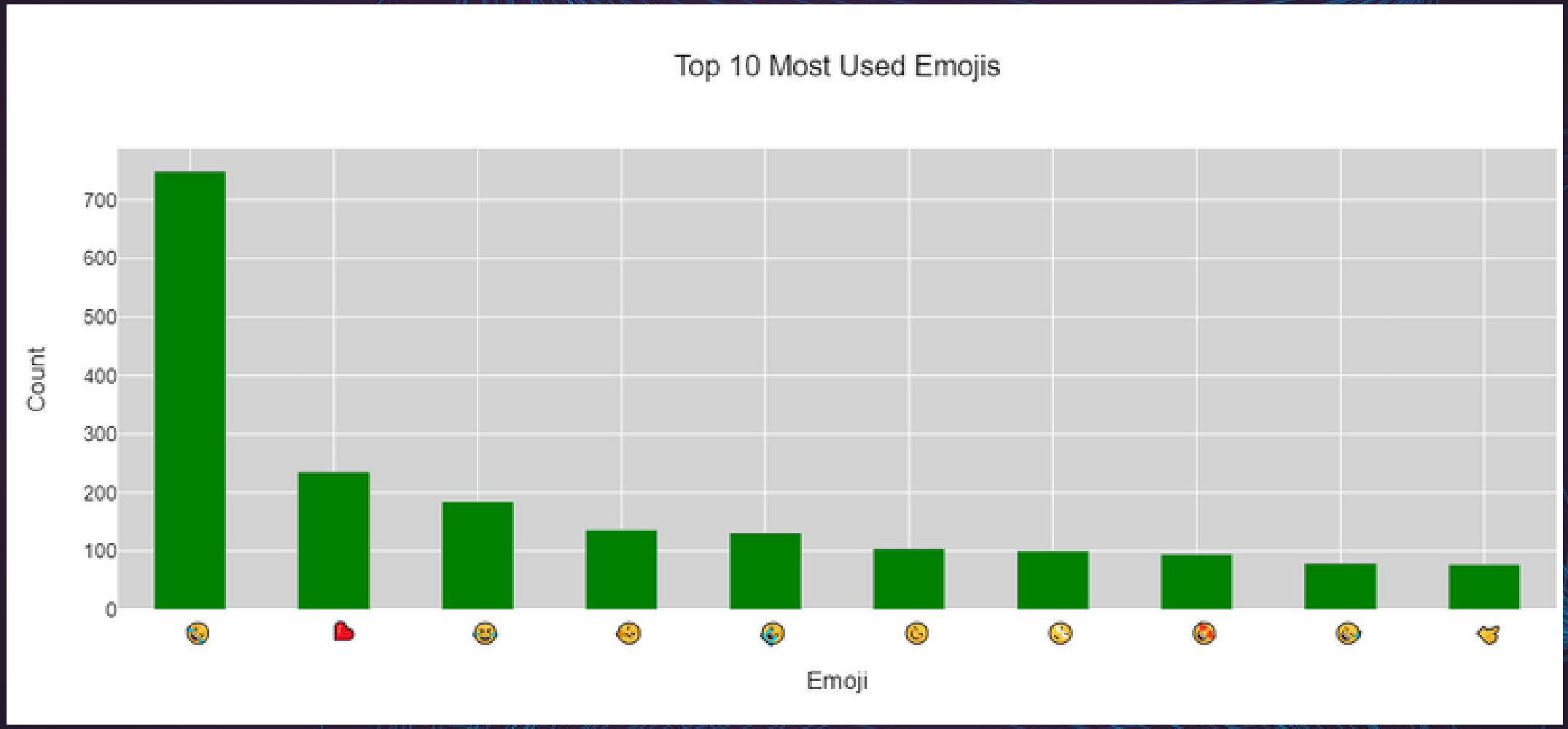
The positive category had the most counts within the dataset while the neutral has the least count.

Positive Emotions Count



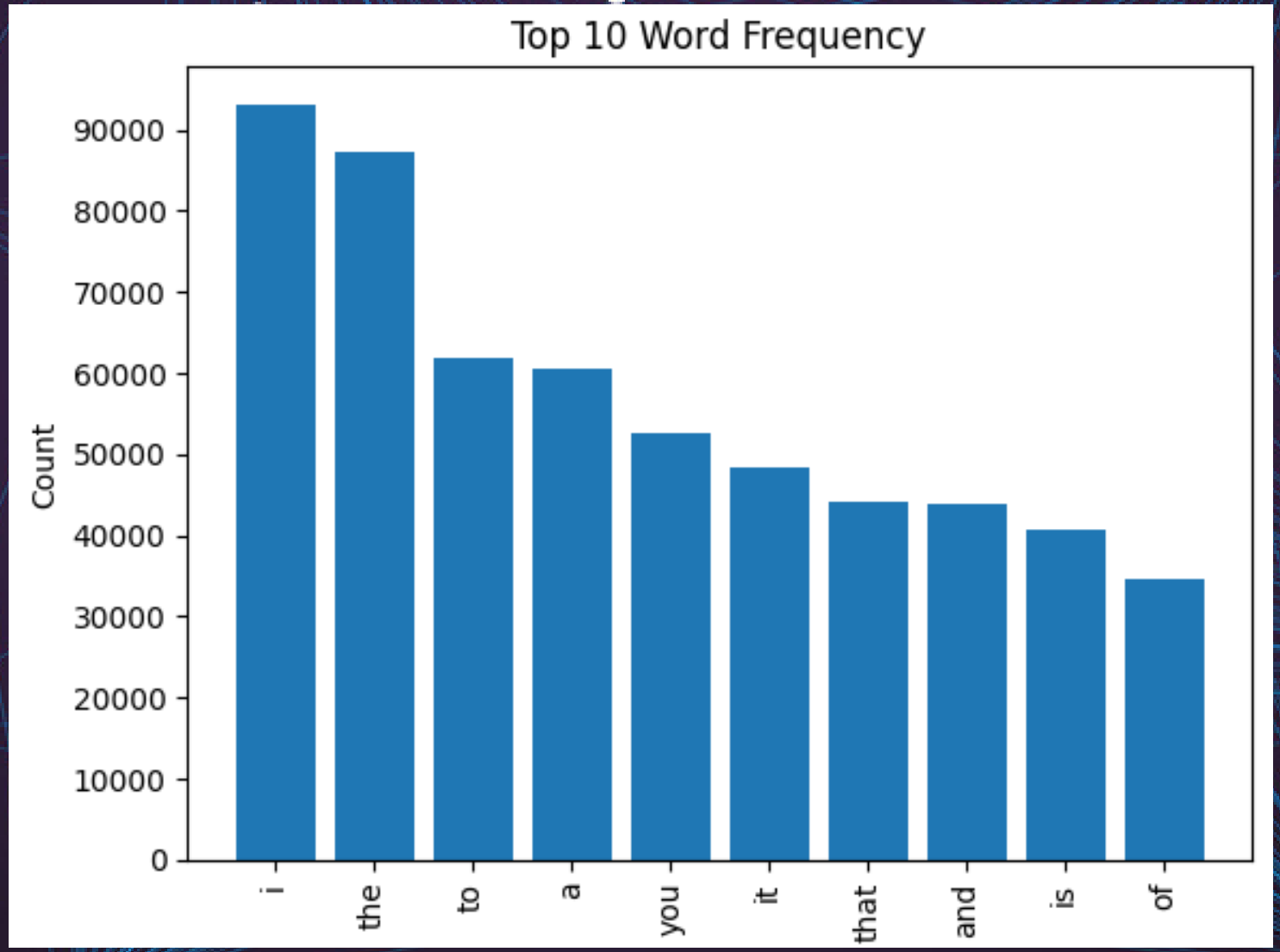
Negative Emotions Count



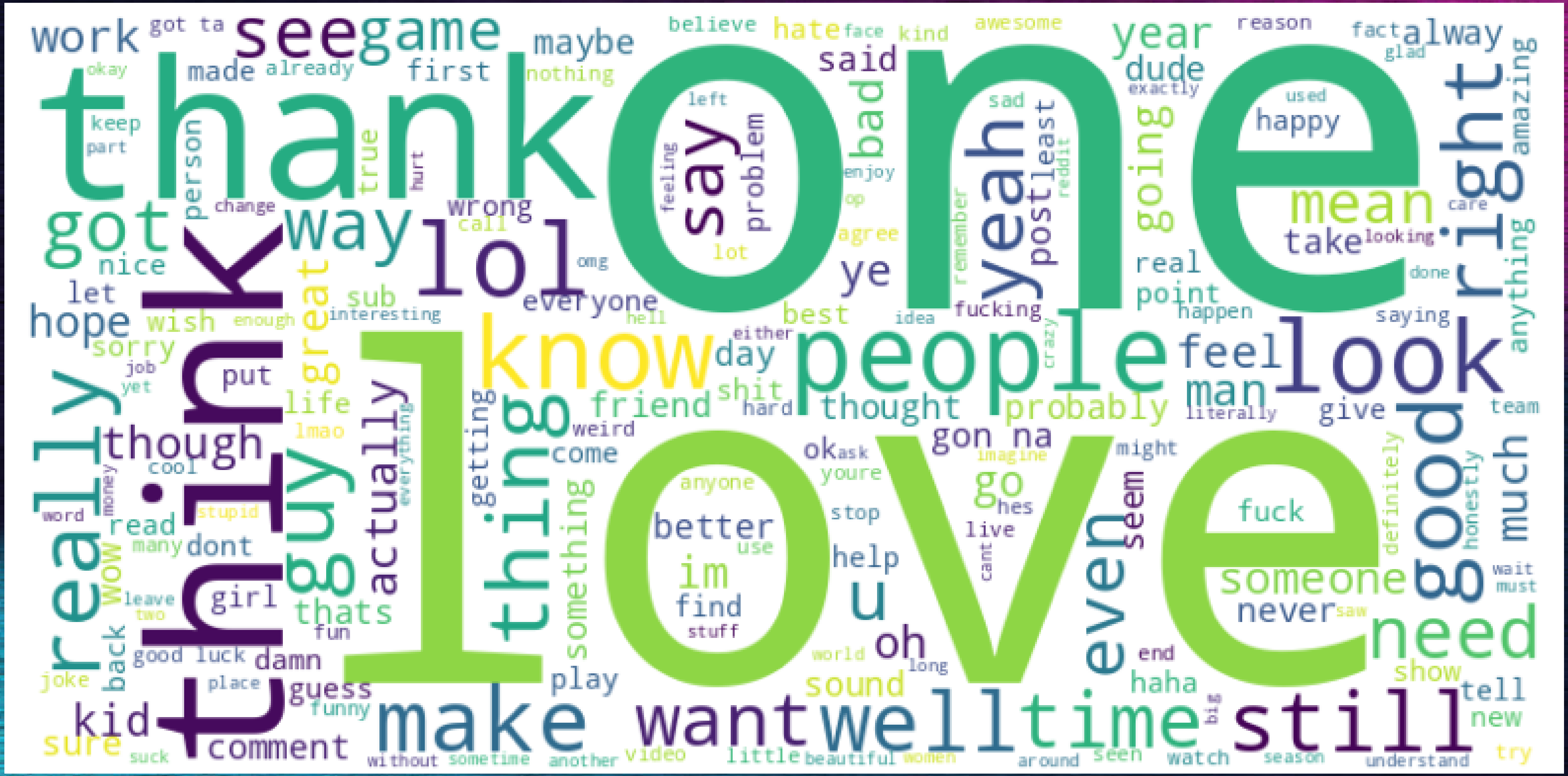


749
235
185
137
132
104
100
95
79
77

Word counts before removal of stop words



Word Cloud



work
got ta
see
game
maybe
believe
hate
face
kind
awesome
year
reason
fact
alway
one
thank
love
think
know
people
look
make
want
well
time
still
really
guy
thing
lol
say
yeah
right
good
got
way
mean
going
dude
happy
amazing
anything
made
already
first
nothing
said
okay
keep
part
person
true
hurt
feeling
left
problem
bad
enjoy
op
sad
post
least
reddit
used
glad
exactly
change
wrong
call
agree
lot
remember
care
take
looking
nice
omg
ye
real
done
let
sub
everyone
point
happen
saying
hope
wish
enough
interesting
hell
best
idea
fucking
either
crazy
sorry
day
feel
job
put
great
shit
yet
life
friend
hard
thought
probably
man
literally
give
team
lmao
weird
gon na
though
getting
come
ok
ask
imagine
might
cool
actually
everything
anyone
youre
go
money
stupid
word
read
better
stop
hes
definitely
much
many
dont
something
live
use
honestly
wow
im
help
cant
fuck
wait
must
leave
girl
find
someone
two
thats
u
seem
even
never
saw
fun
stuff
oh
need
back
good luck
damn
world
long
end
joke
place
play
show
guess
sound
haha
tell
funny
big
kid
new
sure
comment
suck
without
sometime
another
video
little
beautiful
women
around
seen
watch
season
understand
try

Models used:

- Support Vector Machines(SVM)
- RNN (Recurrent Neural Network)
- CNN (Convolutional Neural Network)
- Transformers (Bert)

Models performance comparison

	SVM	RNN	CNN	Transformers
Accuracy	0.96	0.52	0.67	0.62
Recall	0.96	0.52	0.61	-
Precision	0.97	0.52	0.67	-
F1 Score	0.96	0.49	0.64	-

CNN: - This indicated that approximately 58% of all text sentiments were predicted correctly using the CNN model.

RNN: - This indicated that approximately 52% of all text sentiments were predicted correctly using the RNN model.

SVM: - This metric indicates that the model is able to predict the correct sentiment for approximately 96% of the examples in the dataset. This was the highest among the models evaluated.

Conclusions

- Reddit data shows predominantly positive sentiments towards the company.
- The sentiment analysis model using Support Vector Machines demonstrates high accuracy, precision, and recall for positive and neutral sentiments, making it reliable for sentiment analysis.

Recommendations

- For our project, we observed that most of the emotions from the user-generated texts were of a positive sentiment.
- However, we observed low performance metrics on three out of the four models despite cleaning, deduplicating and hyper parameter tuning. Further analysis methods may be performed, and other models may be explored such as Transfer Learning and Word Embeddings.

Challenges Faced

- High computational and time resources required
- Data Preprocessing - the randomness of the text including emojis and text, introduced complexities when preparing the text for modelling
- Data Imbalance - the emotions from the data were skewed, with majority of the sentiments classified as positive emotions. We addressed this by balancing the different sentiment labels namely positive, negative and neutral.