VAIBHAV RASTOGI - CAPSTONE PROJECT

USING AI FOR SENTIMENT ANALYSIS OF CUSTOMER REVIEWS

USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

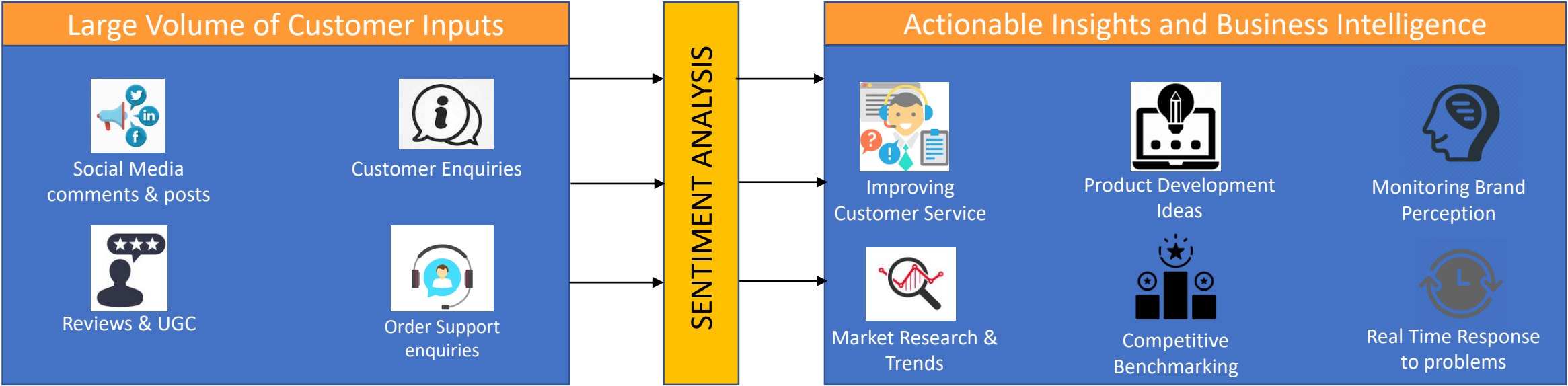> " Sentiment analysis is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention. "

Large Volume of Customer Inputs

- Social Media comments & posts
- Customer Enquiries
- Reviews & UGC
- Order Support enquiries

SENTIMENT ANALYSIS

Actionable Insights and Business Intelligence

- Improving Customer Service
- Product Development Ideas
- Monitoring Brand Perception
- Market Research & Trends
- Competitive Benchmarking
- Real Time Response to problems

DATA

Booking.com customer reviews data – 1 MILLION ++ RECORDS
53% Negative review – 47% Positive Reviews
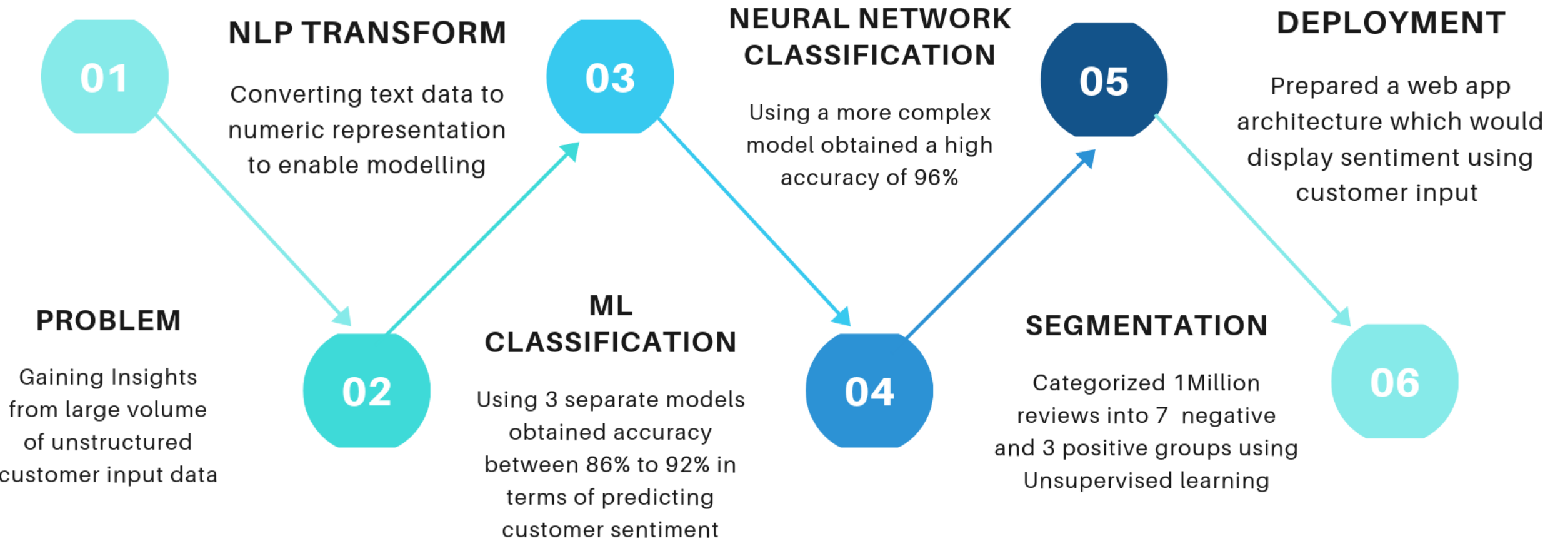19 Million Words – 227 MB Data

Q *How can you mine vast stores of unstructured data to drive actionable insights*
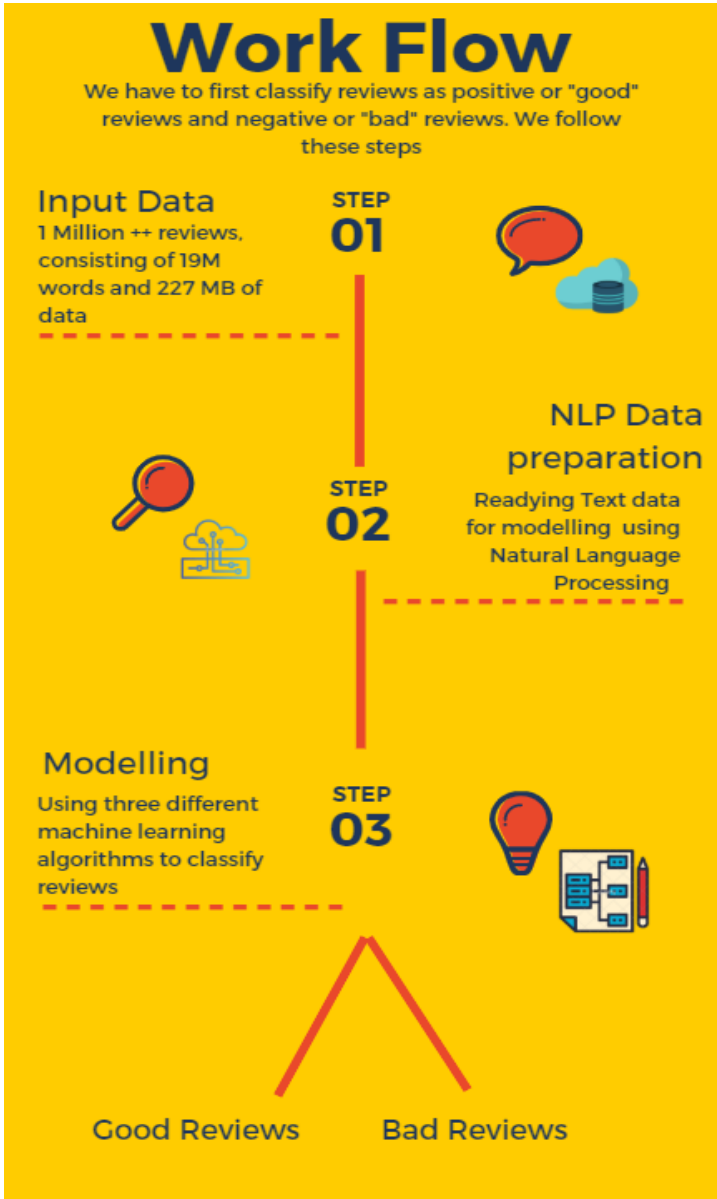
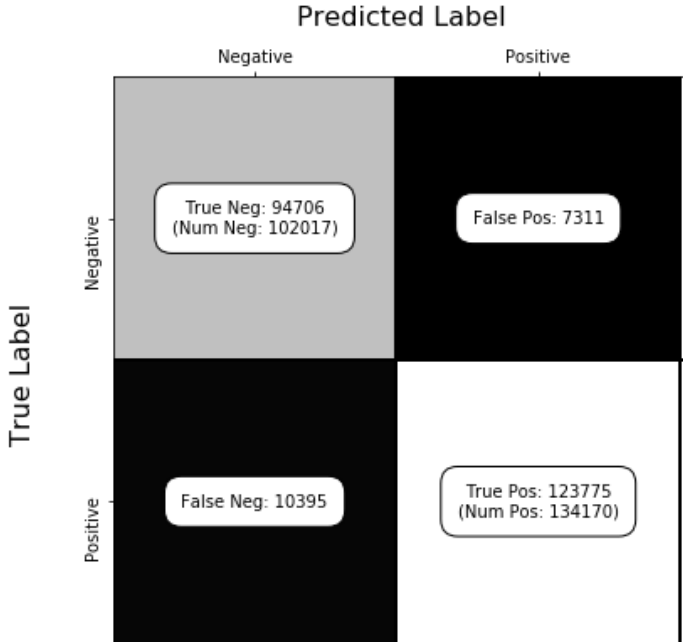# USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

WORKFLOW AND APPROACH

01

PROBLEM

Gaining Insights from large volume of unstructured customer input data

02

NLP TRANSFORM

Converting text data to numeric representation to enable modelling

03

ML CLASSIFICATION

Using 3 separate models obtained accuracy between 86% to 92% in terms of predicting customer sentiment

NEURAL NETWORK CLASSIFICATION

Using a more complex model obtained a high accuracy of 96%

04

05

SEGMENTATION

Categorized 1Million reviews into 7 negative and 3 positive groups using Unsupervised learning

DEPLOYMENT

Prepared a web app architecture which would display sentiment using customer input

06

USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

Work Flow

We have to first classify reviews as positive or "good" reviews and negative or "bad" reviews. We follow these steps

Input Data
STEP 01
1 Million ++ reviews, consisting of 19M words and 227 MB of data

NLP Data preparation
STEP 02
Readying Text data for modelling using Natural Language Processing

Modelling
STEP 03
Using three different machine learning algorithms to classify reviews

Good Reviews   Bad Reviews

Accuracy, Complexity

NLP Methods ->	Uni-grams	Bi-grams	TF-IDF
Modelling Methods ↓			
Logistic Regression	92.2	86.6	92.0
Multi NB	91.6	90.6	91.3
SVM	92.5	86.4	92.4

Confusion Matrix for SVM - unigrams

Predicted Label

Negative   Positive

True Label

Negative

True Neg: 94706 (Num Neg: 102017)

False Pos: 7311

Positive

False Neg: 10395

True Pos: 123775 (Num Pos: 134170)

USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

With ML Classification models we were getting an classification accuracy between 86% to 92%. We will now build a model, a Neural Network and try to get even higher prediction accuracy. A Neural Network "learns" recursively from the error and tries to minimize it with each iteration called an epoch.
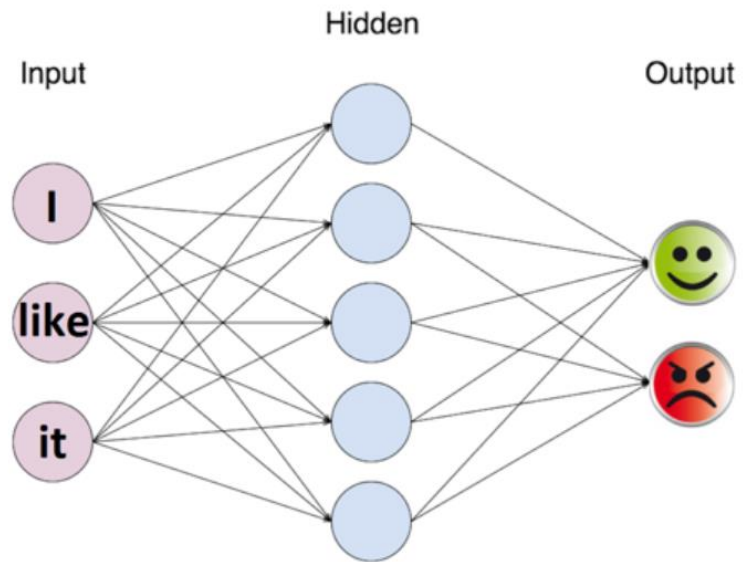
Model Name	Number of Hidden Layers	Hidden Nodes	Train Accuracy	Test Accuracy	Time Taken
Feedforward Neural Network 1	1	10	97.5	96.1	8678 seconds
Feedforward Neural Network 2	1	10	97.2	95.9	215 seconds

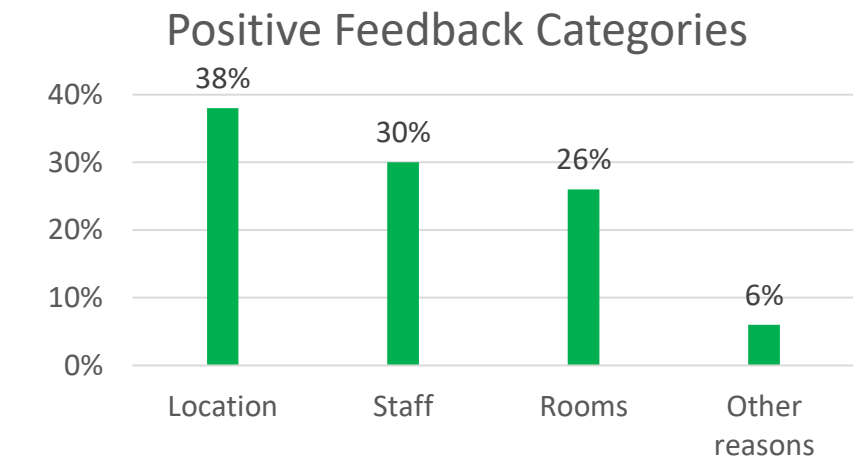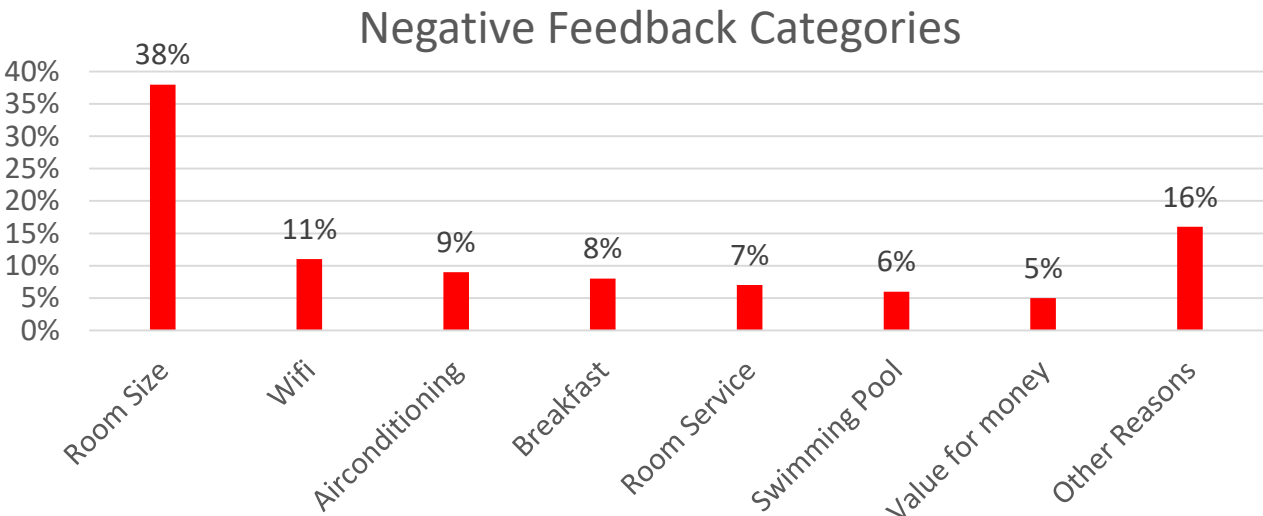Word Positive/Negative Affinity Distribution
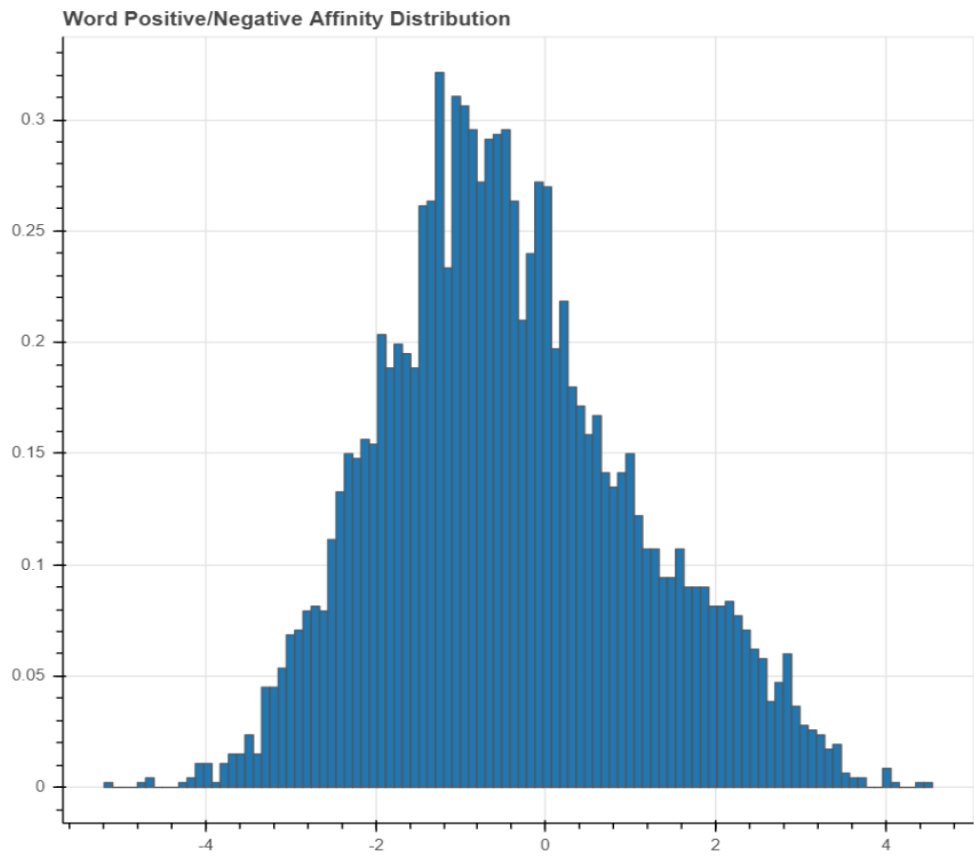
USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

After building predictive models to classify 1M reviews as positive or negative, we segmented them into broader categories
We used TF-IDF and Count-vectorizer NLP and LDA-LSA methods for creating the cluster

Category	Words used	Category	Words used
Room size	Small room, small size, small bed, small bathroom	Room Service	Poor room service, order service, order late
Wifi	poor wifi, wifi didn't work	Swimming Pool	swimming pool small, pool cold, sauna
Aircon	AC not working, hot air, hot room, room cold	Value for Money	poor value for money, expensive
Breakfast	Breakfast Expensive, Small Breakfast, Breakfast poor	Other Reasons	

Category	Words used
Location	Great location, excellent location, near metro, near market
Staff	Staff friendly, staff helpful,
Rooms	Rooms clean, room tidy, room comfortable, bed comfortable
Other reasons	

Negative Feedback Categories

Room Size 38%, Wifi 11%, Airconditioning 9%, Breakfast 8%, Room Service 7%, Swimming Pool 6%, Value for money 5%, Other Reasons 16%

Positive Feedback Categories

Location 38%, Staff 30%, Rooms 26%, Other reasons 6%

USING AI FOR SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS

"Deployment to production is a method that integrates a machine learning model into an existing production environment so that the model can be used to make decisions or predictions based upon data input into the model."
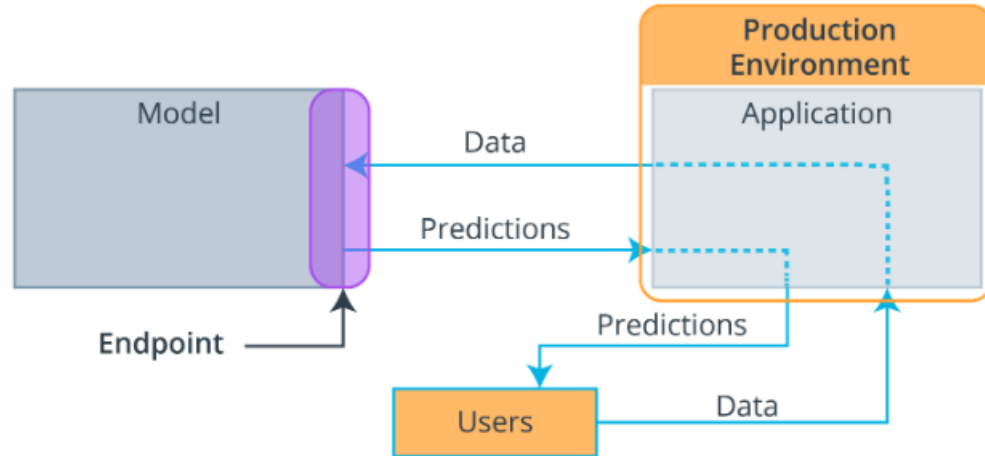
ARCHITECTURE

- This architecture has been replicated on Amazon Web Services Sagemaker
- The prediction is the lambda functions, the data is sent through an API
- The model is our sagemaker notebook and the app is a web app

WEB APP

This simple web app sends information to the model using API and receives information from the model using the lamba function

Is your review positive, or negative?

Enter your review below and click submit to find out...

Review:

I had a terrible experience at the hotel, my room was very uncomfortable because of ineffective cooling

Submit

Your review was NEGATIVE!

QUESTIONS?

Appendix Slides