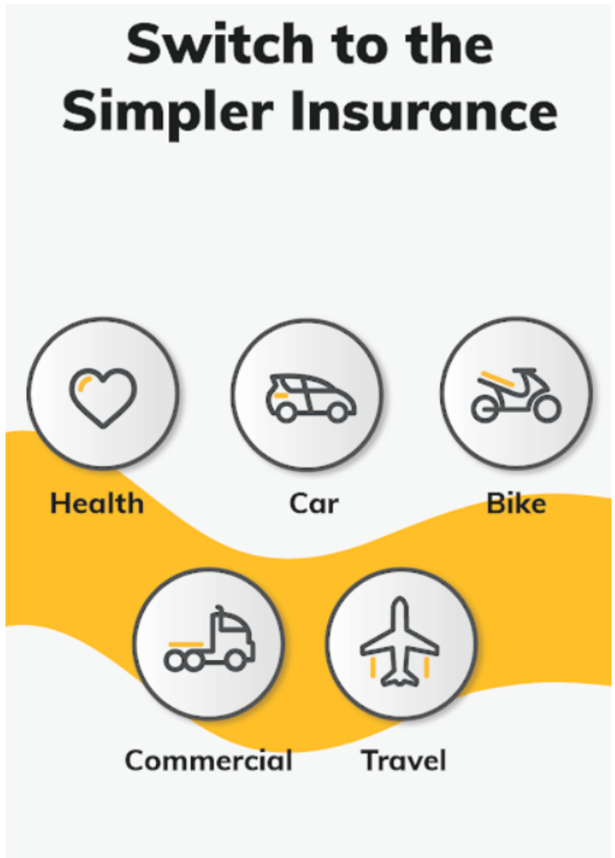
# Sentiment Analysis of Insurance Apps Reviews




Source : Digit Insurance

Clear and crisp instructions are given .easy to claim.very much compact but informative.quick approval obtained each time.App is too good to work with

Waste application I have seen ever. We can't renew our policy also app is dead slow every time.


Rohan Prabhakar Agale
Springboard Capstone 3

Introduction

Data

1. Data Wrangling

2. Exploratory Data Analysis
    - 2.1 Distribution of review length
    - 2.2 Negative Review Trends by company & year
    - 2.3 Most frequent words by sentiment
    - 2.4 Most frequent words by sentiment

3. Modelling
    - 3.1 Evaluation matrix
    - 3.2 Modelling strategy
        - 3.2.1 Modelling :
    - 3.3 Best Model

4. Business Insights

# Introduction

In this digital era, Mobile applications offer a convenient and seamless channel to buy and renew insurance policies. Customers can quickly and easily lodge claims and get prompt service. Applications are effective communication tools for consultations and notifications. With these applications, companies can build loyalty, increase customer engagement and offer customized products. Hence, it is crucial for companies to build an application that lives up to these expectations.

Almost all General Insurance(GI) companies have a mobile application on Play store. We will leverage Play Store reviews to **analyze customers' feedback to know what is working, what are the challenges and how companies can address those challenges**.

Before we start, let us understand what performance & features the app should offer.

**Must haves :**
1. Simple & easy to operate
2. Fast to load & navigate
3. Quick to buy/ renew policies
4. Instant claim filing & tracking
5. View and edit policy or claims details
6. In-app customer support using live chat/call

These are the bare minimum requirements for an app to have a smooth & satisfactory experience of core insurance products.

**Good to have :**
1. List of Network providers - Easy access to list of hospitals or garages.
2. Emergency assistance - Feature to connect to the ambulance/nearest hospital for health insurance & garage for auto insurance.
3. Self-inspection using smart phone for filing claim or renewing break-in policy
4. Tele-consultations with experts like Doctors, Dieticians etc
5. Notifications about renewals, claims tracking, preventive tips & weather events.

# Data

We will scrape reviews from Google Play Store. We can identify the company using a unique id on the play store as highlighted below.

https://play.google.com/store/apps/details?id=com.ba.cp.controller&showAllReviews=true

Using **google_play_scraper** api, we will one by one extract reviews for all the companies. Below is a sample of 5 reviews.

| reviewId | content | score | thumbsUpCount | reviewCreatedVersion | at | replyContent | repliedAt | appId |
|---|---|---|---|---|---|---|---|---|
| k3baCjf8... | Excellent.. The service and reaction were exce... | 5 | 0 | NaN | 2021-08-26 22:29:04 | NaN | NaN | icici.lombard.ghi |
| :ILU1eSsx-C2pCW6... | The app is simple and tidy, with a focus on th... | 5 | 0 | NaN | 2021-08-26 22:28:35 | NaN | NaN | icici.lombard.ghi |
| BJSkmZ... | It is simple to select and purchase insurance.... | 5 | 0 | NaN | 2021-08-26 22:27:21 | NaN | NaN | icici.lombard.ghi |
| AhMokel... | I've used this app for my car insurance a few ... | 5 | 0 | NaN | 2021-08-26 22:26:59 | NaN | NaN | icici.lombard.ghi |
| 18D2C14... | It is wrost insurance company . One of my frie... | 1 | 0 | 2.0.40 | 2021-08-26 21:21:10 | NaN | NaN | icici.lombard.ghi |

For our analysis, we are mainly interested in below fields

**1. reviewId -** unique identifier for reviews
**2. content -** actual review text
**3. score -** customer rating between 1-5
**4. at -** time of review posting
**5. appId -** unique identifier of the company
* * The data is scraped from the Play Store as at 26 Aug 2021

# 1. Data Wrangling

In Data Wrangling, we explore the data for
1. **Completeness** : Each record contains all the variables and records are unique.
2. **Validity** : We will be focusing on reviews in the English language.
3. **Credibility** : Do we have relevant and enough information in the review text ?
4. **Cleaning**: Accordingly, we will perform typical data cleaning steps.

Below are our findings from the Data Wrangling exercise.
**1. reviewId -** observations are unique by reviewId and there are no missing values
**2. content -** We have 6 reviews without any content. **These 6 records are deleted**.
**3. score -** customer rating are complete and are within the valid range of 1-5
**4. at -** dates are complete and are within the reasonable range.
**5. appId -** unique identifier of the company is complete.

For Our Analysis purpose, we will consider the top 5 private GI companies by Gross Premium in 2020-2021 & 3 insurtech companies. These companies contribute to 93% of total reviews.

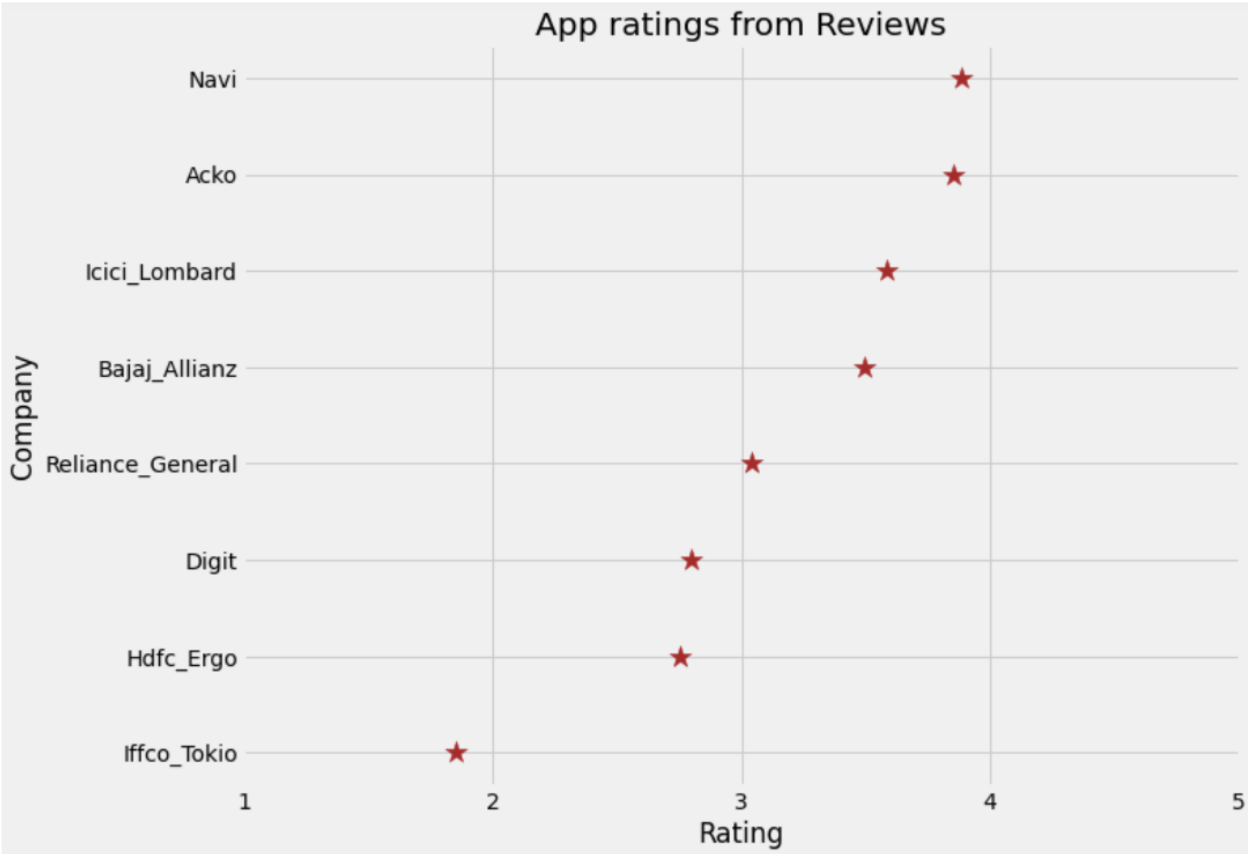
*top_5 = ['icici_lombard', 'bajaj_allianz', 'hdfc_ergo', 'iffco_tokio', 'reliance_general']*
*insure_techs = ['acko', 'digit', 'navi']*
So we have **28981 reviews** for analysis. Let us look at these reviews by years.


.

We have just 404 reviews before 2016. Customers' feedback about applications really picked up in 2018. In 2019, the number reached a respectable level of 3800 reviews. Growth continues in 2020 & 2021. **We will not consider the reviews prior to & including 2018 year** because

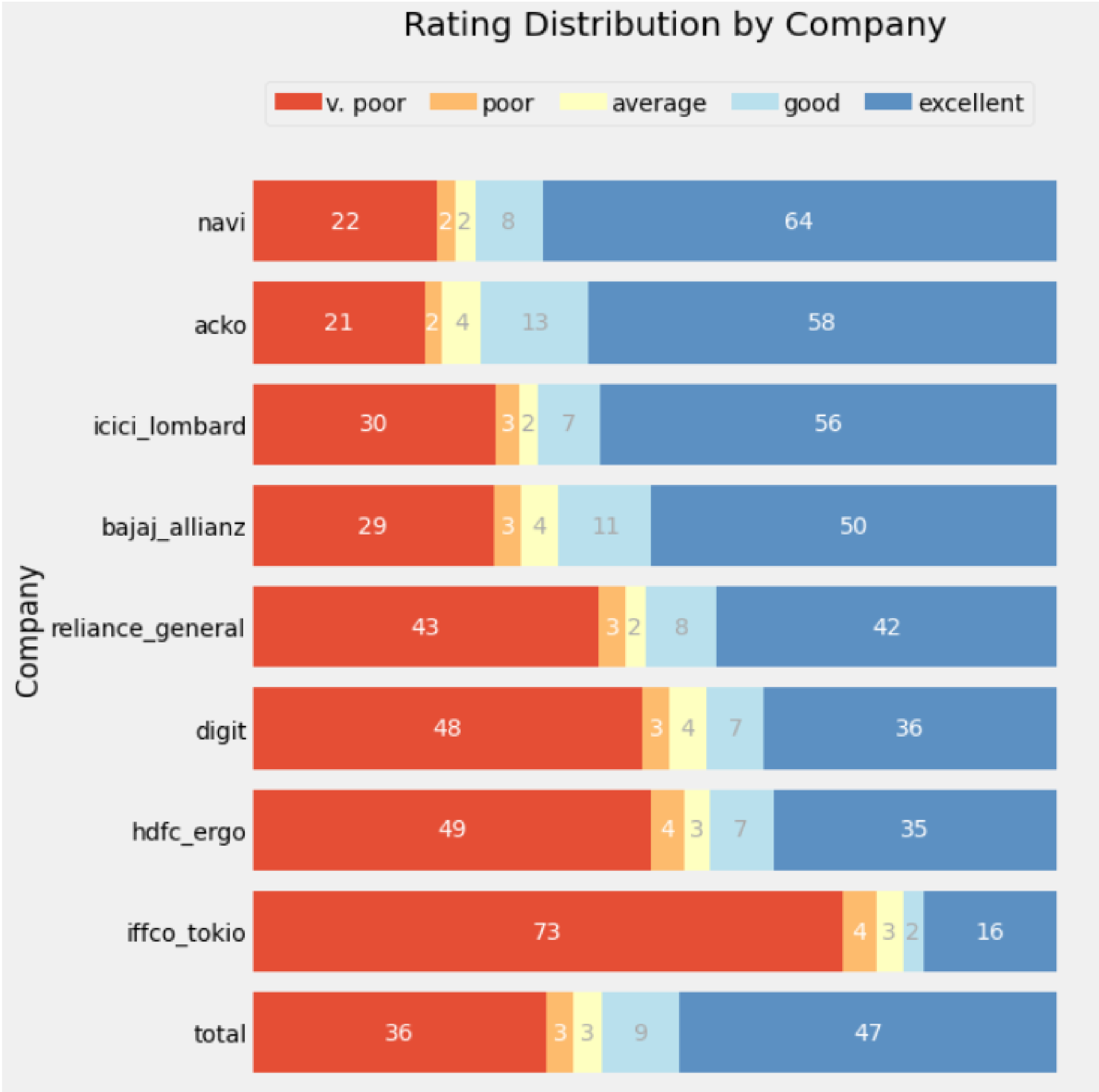
1. Numbers are less.
2. Mobile application improvements happen much faster.
3. Those are initial development and adaptation years and may be lacking in maturity.

** 2021 is as of 26 Aug 2021.




Navi & Acko are leaders with ~3.8 stars. ICIC Lombard & Bajaj Allianz score ~3.5 stars. These GI companies are adopting & making good progress in the mobile application channel. Digit & HDFC Ergo are at 2.8 stars & need to catch up with the leaders. IFFCO Tokio is a laggard with just 1.9 stars.

Let's look at the distribution of ratings.




** Numbers indicate % of ratings for the score.

Navi has highest i.e. 64% 5 star reviews. Acko & ICICI Lombard are next with 58% & 56% 5 star ratings. Yet these companies have a 20-30% 1 star rating for apps. This indicates a huge scope of improvement even for leaders. IFFCO Tokio has a 73% 1 star rating and is an outlier. Digit, HDFC Ergo & Reliance have 40-50% reviews with just 1 star rating.

Overall, we have a highly polarized dataset. ~80% of the reviewers have given either Excellent or Very Poor rating. Only about 3% reviewers are neutral. We group the ratings as per below logic

1. combine 4 & 5 as positive and label them as 0

2. combine 1 & 2 as negative and label them as 1

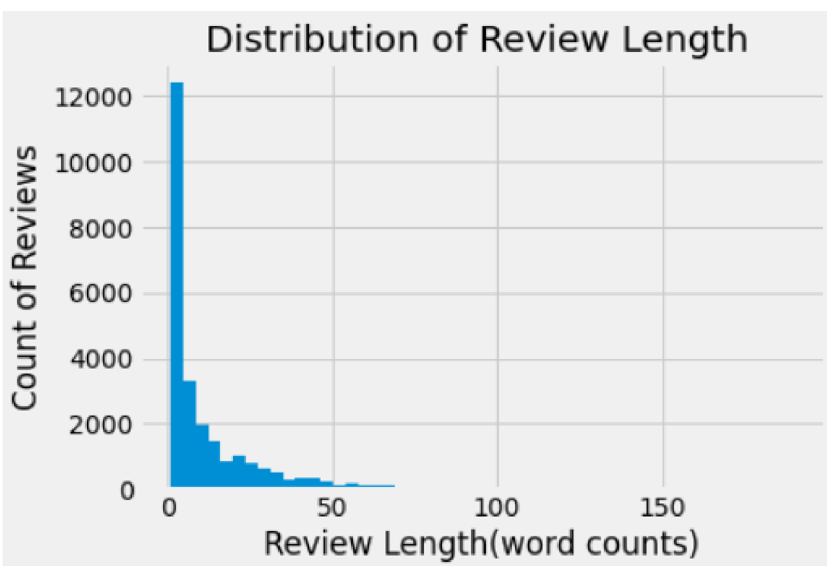
3. ignore 3 (931 reviews)

We will call this as **rating** variable. We have below distribution of reviews by sentiment & company

| sentiment | counts |
|---|---|
| Positive | 15010 |
| Negative | 10558 |

| Company | Reviews |
|---|---|
| hdfc_ergo | 6882 |
| bajaj_allianz | 5949 |
| icici_lombard | 4523 |
| acko | 4016 |
| reliance_general | 2692 |
| iffco_tokio | 665 |
| navi | 447 |
| digit | 394 |

We plot the distribution of review text by word count.




We have a highly skewed distribution. 50% of the reviews are having less than 5 words while top 5 percentile reviews are greater than 54 words in length. Longest review has 190 words.

Let us look at reviews with short words

**Reviews with 1 word :**

```
['good' 'pathetic' 'Perfect' 'Good' 'Good' 'Good' 'Excellent' 'Aws' 'Nice'
'good' 'Nice' 'Awesome' 'Nice' 'good' 'Nice' 'good' 'Nice' 'Excellent'
'Crash' 'Best' 'Seamless' 'Good' 'good' 'super' 'Good' 'Good' 'Great'
'Super' 'Good' 'Good']
```

**Reviews with 2 word :**

```
['V good' 'Safe app' 'Good app' 'Very convenient' 'Very helpful'
'Nice App' 'Nice app' 'GOOD support' 'Bad experience' 'Nice app'
'Great app' 'Nice app' 'good application' 'Not responding' 'Rohit Maweba'
'Nice app' 'No comment' 'great app' 'Very good' 'Very bad' 'Usable app'
'Good app' 'NICE APP' 'Very Good' 'Good response' 'Horrible App'
'nice services' 'Super 👌' 'Nice service' 'site crash']
```

**Reviews with 3 word :**

```
['Very slow app' 'datta, kale' 'Very nice app' 'not user friendly'
'VERY GOOD APP' 'Very goood service' 'Easy and fast'
'Very Useful friendly' 'Too much bugs' 'Simple and easy' 'Super & Easy'
"Doesn't work" 'Very nice application' 'Not getting otp'
'Worst experience ever' 'Bad very bad' 'Very good experience'
'Good Dr. sarvice' 'Awesome service...' 'Good application Nhea'
'Instant and fast' 'very good app' 'very good service'
'Easy to processes' 'need to evaluate' 'Cheater company......'
'Committed to customer' 'Super Very eazy' 'Made process easy'
'amazing service..']
```

**Reviews with 4 word :**

```
['after update not working' 'It is not downloading'
'Useless .advisor are lier' 'Best app for insurance'
'Very bad work slowed' 'Very very bed service' 'Best and good 👌'
'Waste insurer worst app' 'Very very bad app' 'It is not instolde'
'Nice user friendly app' 'Will never work properly' 'Very very bad app'
'Reasonably good working superb' "App doesn't work"
'Best service given me' "IT'S REALLY FAST" 'Help for tracking policy'
'Hassle free insurance renewal' 'Easy to Use.']
```

Reviews with less than 4 words are largely nothing but ratings. Eg  Good, Excellent, Nice App, Bad Experience etc. They are not adding any specific/generic information about the apps.

| Company | Long Reviews | Reviews |
|---|---|---|
| hdfc_ergo | 4084 | 6882 |
| icici_lombard | 2957 | 4523 |
| bajaj_allianz | 2762 | 5949 |
| acko | 2011 | 4016 |
| reliance_general | 1535 | 2692 |
| iffco_tokio | 491 | 665 |
| digit | 273 | 394 |
| navi | 180 | 447 |

Because of a lack of credible data, we will do the analysis for below 4 companies.

*['icici_lombard',  'bajaj_allianz',  'hdfc_ergo',  'acko']*

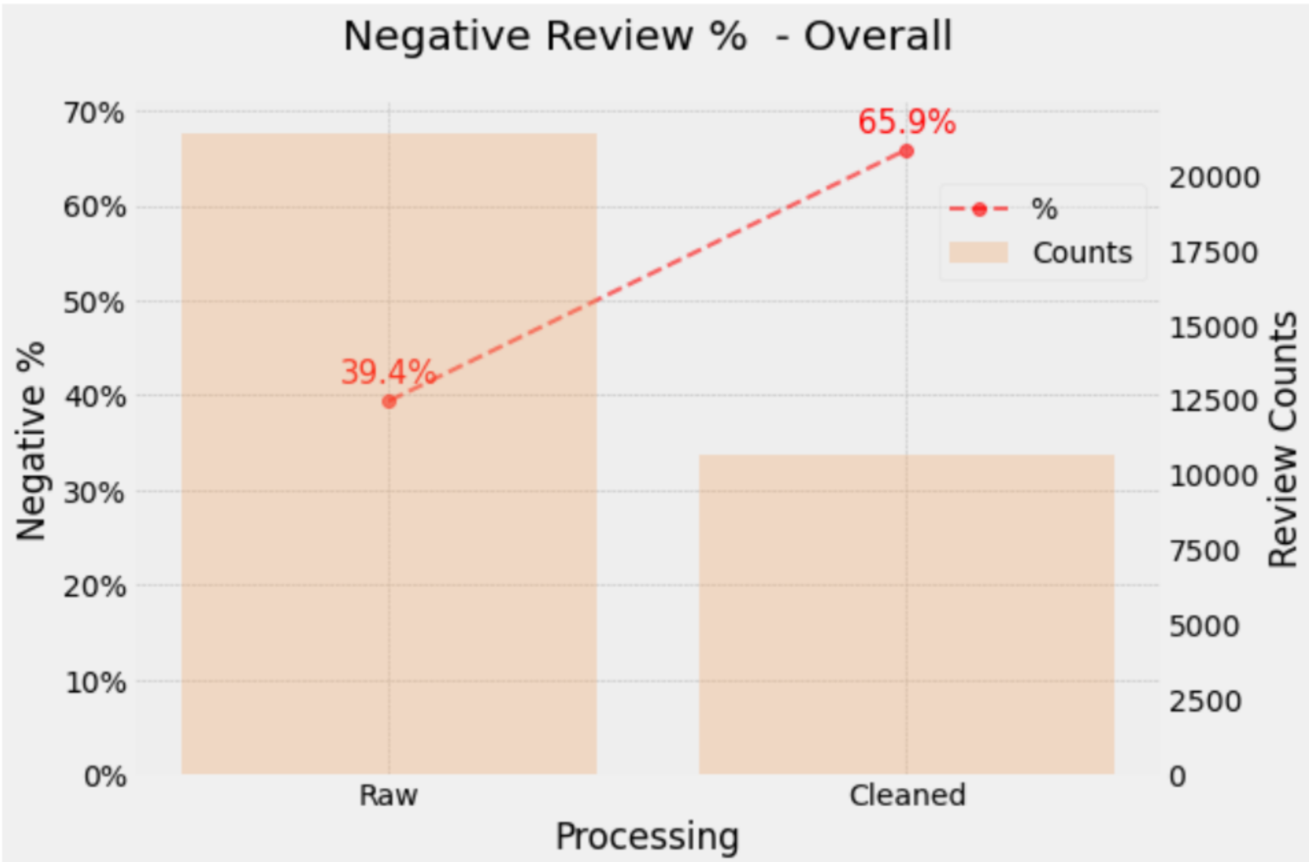
Now we will perform the typical text data cleaning steps.

1. Remove leading and trailing whitespaces
2. Expand contractions.  Eg I've  is converted to I have
3. Convert emojis to text
4. Convert emoticons to text
5. Convert text to lower cases
6. Remove non-alphanumeric text. This will remove reviews in Indian languages
7. Remove brand names. Eg ICICI Lombard, Acko etc
8. Remove extra white spaces.

Once above steps are complete, we will

1. Lemmatize the text
2. Removed stop words

Again we removed reviews with length 1 & 2 from clean text data for the same reasons. Here is a comparison between Raw data & clean data.




We have deleted ~50% of the reviews (from 21370 to 10640). Most of the deleted reviews were of positive sentiments. This makes sense as many customers with positive experience would write short reviews with generic words. Hence, these reviews are deleted in data cleaning. The net effect is that% of Negative ratings increase by 67%.

This has essentially reversed the majority class from positive to negative sentiment. Our main objective is to find what is driving the negative sentiments. Hence, it is OK if the probability of negative reviews changes. The positive reviews which are removed are sparse and do not contain useful information.

We observe the same effect of increase in Negative ratings ranging from 48% to 100% across different companies.

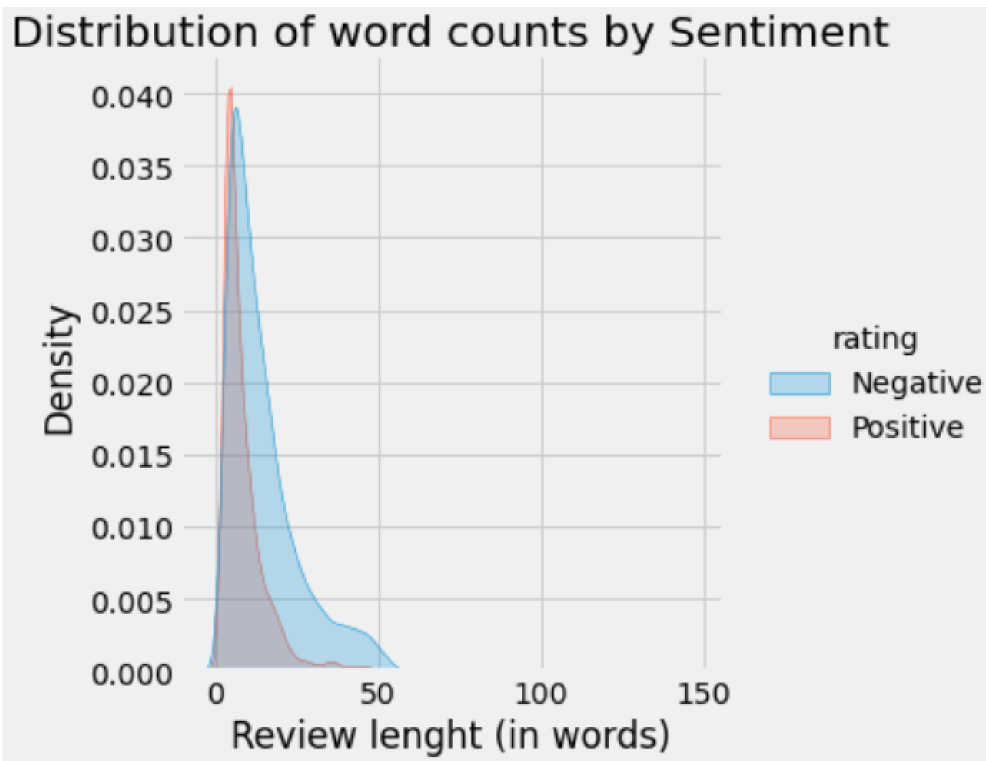
# 2. Exploratory Data Analysis

First, we will split the data using stratified sampling on

1. rating
2. company

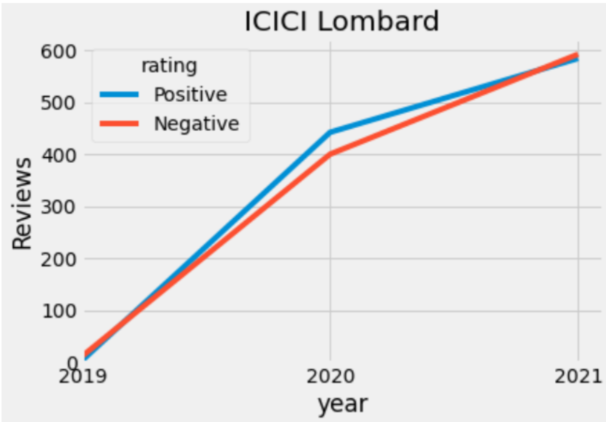
to ensure a representative test set. The size of train data is 7980 reviews & test data is 2662 (25%). EDA is carried out on train data.

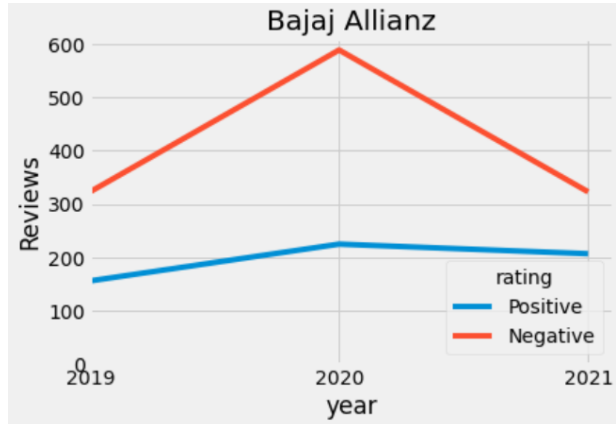
## 2.1 Distribution of review length

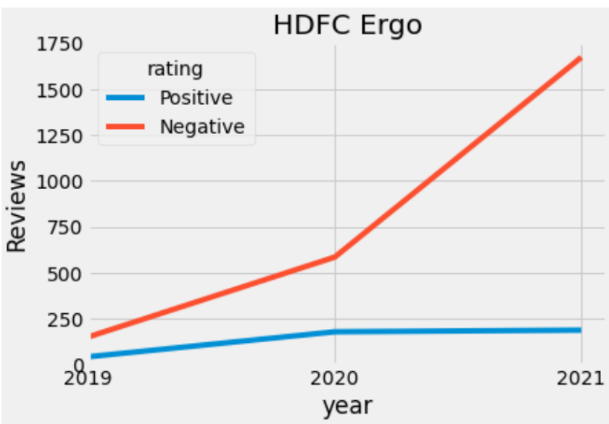


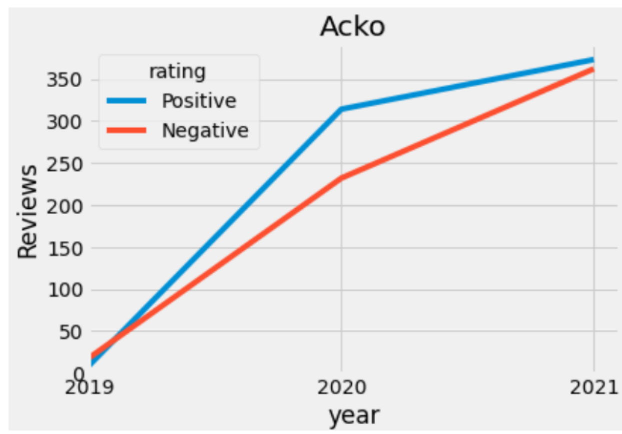

As we had noted before, on an average, positive reviews are shorter in length. Distribution dissimilarity between positive & negative reviews is greatly reduced.

## 2.2 Negative Review Trends by company & year



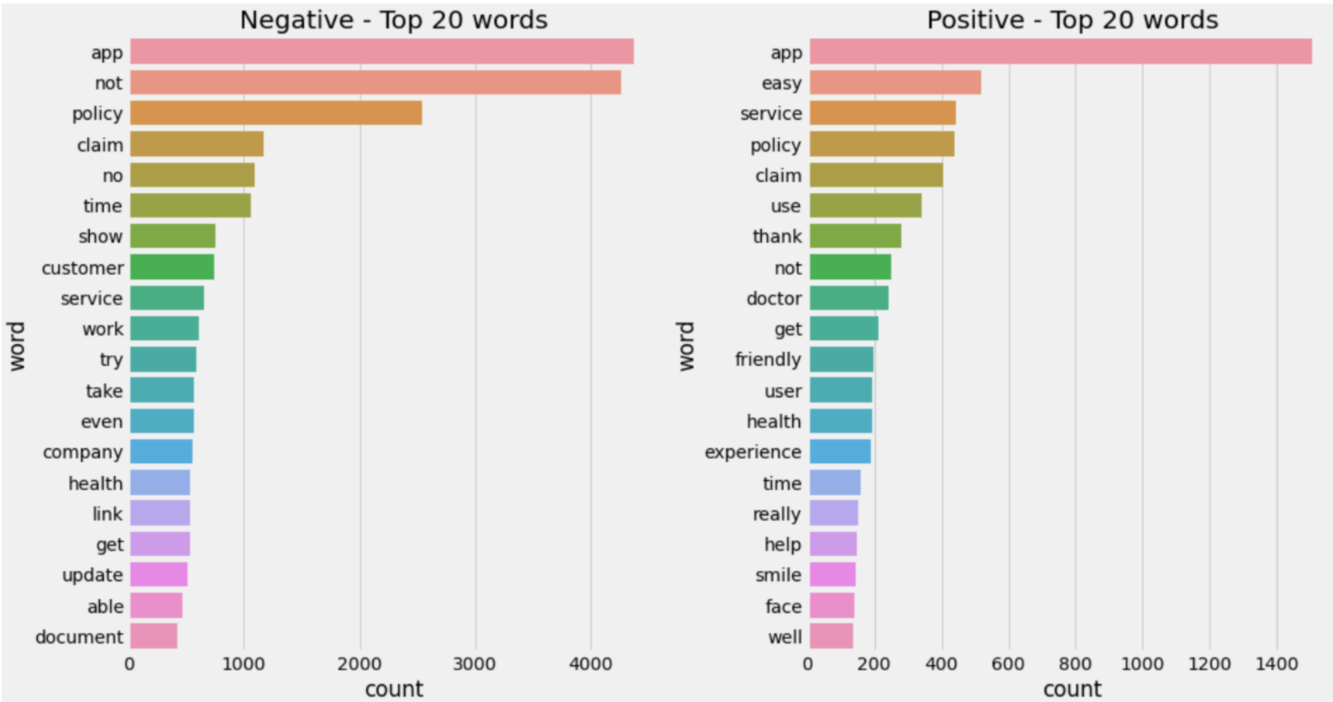




Bajaj Allianz has significantly brought down negative reviews by half in 2021. While HDFC Ergo has a clear problem of almost 3 times increase in negative reviews in 2021. For ICICI Lombard & Acko, growth in negative reviews is in line with trend.

## 2.3 Most frequent words by sentiment



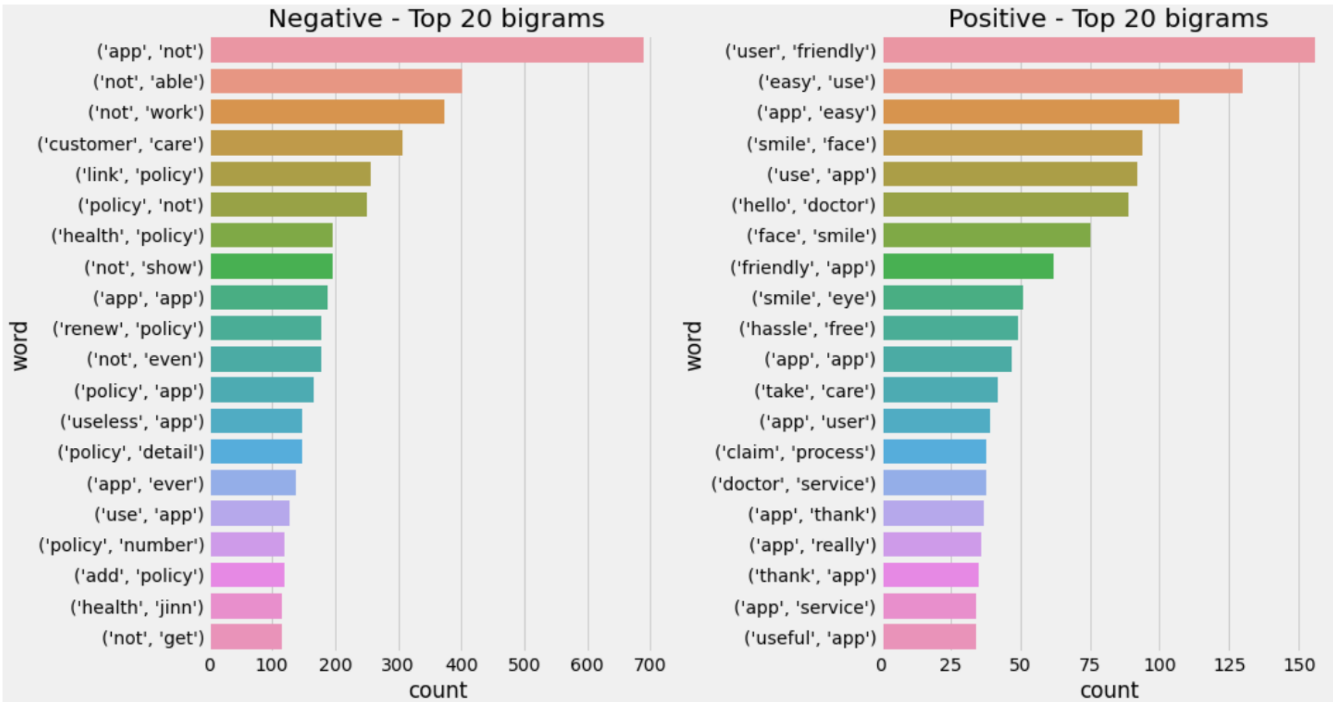

If we ignore app, not, and no, then we have below __negative__ top frequent words :

1. Policy
2. Claim
3. Time
4. Show
5. Customer/Service

If we ignore app, use and thank like generic words, then we have below __positive__ top frequent words :

1. Easy
2. Service
3. Policy
4. Claim
5. Doctor

Policy, Claims & Service are present in both the sentiments. Let us look at bigram & trigram for better understanding.




Based on top frequent words, biagrams & trigrams,
following factors are driving negative sentiments

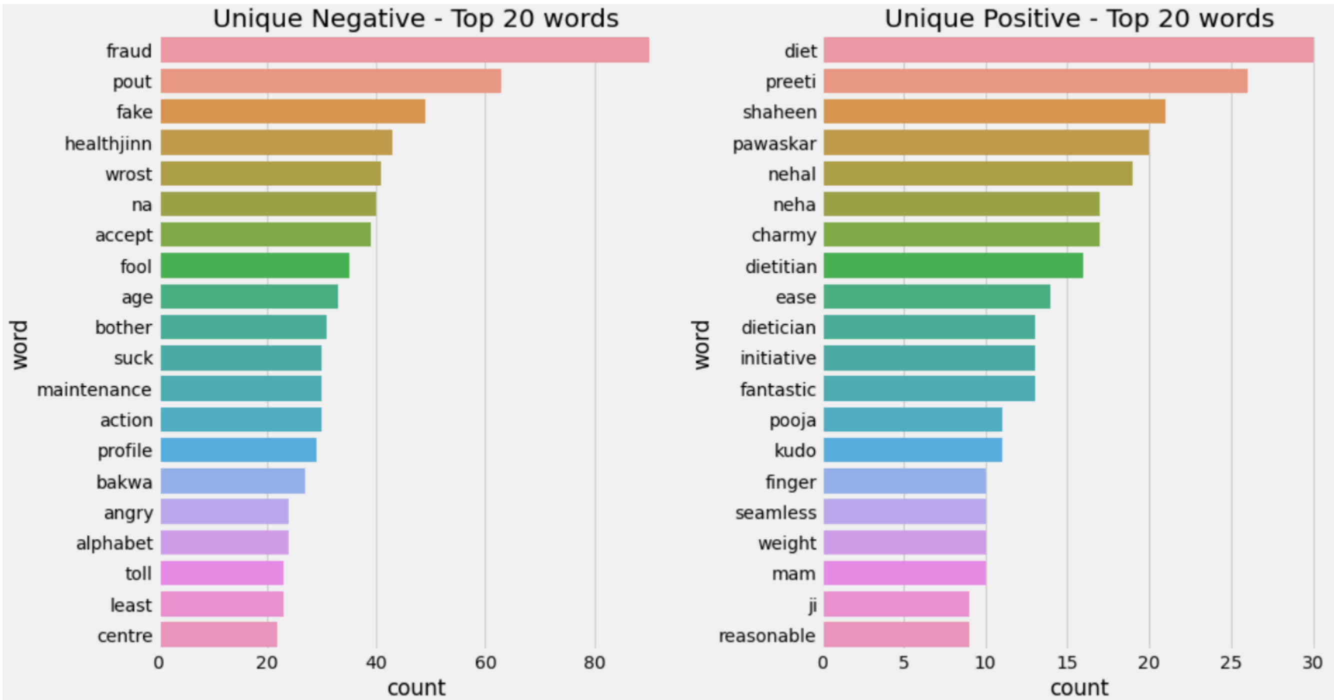
1. App is not functioning as per expectation.
2. App is slow.
3. Some customers are not able to login as well.
4. Customers are not able to link/renew policy.
5. Customers are not happy with the response of customer care to their queries.

And following factors are driving positive sentiments

1. Ease of use
2. Hello Doctor service/feature
3. Claims process

Hello Doctor is a tele-consultation service for health issues. It is provided by ICICI Lombard. We can clearly see that this value added service is well operational and customers are benefitting from it.
source :
https://www.outlookindia.com/outlookmoney/insurance/takecare-app-a-hit-with-customers-says-icici-lombard-7120
On the other hand, Health Jinn was an app from Apollo Munich company which was later acquired by HDFC Ergo in Jan 2020. Customers were not happy with this app.
source : https://www.apollomunichinsurance.com/miscellanous/download-health-jinn.aspx

## 2.4 Most frequent words by sentiment




Negative reviews contain generic words except Health Jinn. We see that diet consultation service is specifically mentioned in the positive review. This service is a part of Hello Doctor feature by ICICI Lombard. We also see names like preeti, shaheen etc. mentioned frequently in reviews. These are the names of either Dietician or customer executive of ICICI Lombard.

1. Dieticians are helping customers on diet plan through online chat
2. Customer executives helped customers in installing the app and explained the app features.

This strategy of increasing app adaptation and providing value added services seems to be working well for the company and customers.

Company wise trends in frequent words are almost the same as overall trends with few exceptions.

1. ICICI Lombard - Hello Doctor feature is the top positive frequent word and specific to ICICI Lombard.
2. Bajaj Allianz - Customer Service is worse than app related issues.

# 3. Modelling

In this step, we will define an appropriate evaluation matrix, devise modelling and validation strategy, select the best model and share actionable business insights.

## 3.1 Evaluation matrix

For the GI industry, it is important to increase the adaptation and usage of mobile applications. We are more interested in understanding what kind of application features or customer services are driving positive or negative sentiments. We have seen that 40% of customers are not happy with mobile apps. This is a huge number. Hence, our primary focus will be negative sentiment. Given this business objective, we will be using the **pr_auc matrix** as it gives more importance to positive class which in our case is negative sentiment.

## 3.2 Modelling strategy

We are more interested in interpretation and insights. Hence, we consider following algorithms for our classification task.

### 3.2.1 Modelling :

1. **Logistic Regression with l1 regularization** - Logistic regression is a simple linear algorithm with decent performance. We will use l1 regularization for feature selection. This will also be our base model.
2. **MultinomialNB Bayes** - MultinomialNB Bayes is well suited for text classification tasks with discrete features.
3. **XgBoost** - XgBoost is known to outperform linear models in many competitions.

### 3.2.2 Validation :

We will use a **3-fold cross-validation** strategy on train data for hyperparameter tuning and model selection. We will train these models on the same set train-validation datasets. To select the best hyperparameters, we will use **RandomSearchCV** technique.And finally, we will **test the performance of the best model on the test dataset**. This will give us a better idea of how well the model performs on unseen data.

### 3.2.3 Feature Extraction :

For feature extraction from text, we will use CountVectorizer & TfIdf Vectorizer. We will first find the suitable minimum document frequency(min_df) using logistic regression. Best min_df is 80 for count vectorizer & 160 for TfIdf vectorizer.

## 3.3 Best Model

| Best Model from RandomSearchCV | Vectorizer | Iterations | Mean roc_auc score | Standard deviation of roc_auc score |
|---|---|---|---|---|
| Logistic Regression | Count | 20 | 0.941243 | 0.003511 |
| Logistic Regression | TfIdf | 20 | 0.938307 | 0.001265 |
| Multinomial Naive Bayes | Count | 20 | 0.940180 | 0.003642 |
| XgBoost | TfIdf | 20 | 0.938992 | 0.000500 |

Logistic Regression with count vectorizer has a pr_auc of 0.94124. While MultinomialNB with count vectorizer has a pr_auc of 0.94018. Both models perform well and are within 1 standard deviation of each other.

We will **choose MultinomialNB** because it is **directly interpretable** and the contribution of each feature towards the negative class is very clear. Also, conditional probabilities are relatively **easy to explain to business** people than Logistic Regression Coefficients. Now we see how well MultinomialNB performs on unseen test data

Performance on Test Data :

**0.94440** score on test data tells that our model generalizes well on test data. This score is slightly greater than 0.94018 cross validation score but is within 1.2 std deviation(0.0036) from it. This confirms that our model generalizes well on unseen data.

Let us look at the top 15 most important features by sentiment.

**Negative Sentiment**

| words | P(Negative \| word) |
|---|---|
| slow app | 0.999784 |
| loading | 0.984085 |
| work properly | 0.966777 |
| exist | 0.966385 |
| every time | 0.957028 |
| show policy | 0.950579 |
| add policy | 0.947608 |
| load | 0.945043 |
| login | 0.943337 |
| nothing | 0.940099 |
| open | 0.936334 |
| ask | 0.933147 |
| many time | 0.931878 |
| never | 0.929929 |
| properly | 0.927458 |

**Positive Sentiment**

| words | P(Negative \| word) |
|---|---|
| smile face | 0.000849 |
| smile face smile | 0.001108 |
| smile face smile eye | 0.001660 |
| doctor service | 0.002041 |
| app thank | 0.002339 |
| face smile | 0.019022 |
| smile eye | 0.027706 |
| face smile eye | 0.027706 |
| app easy use | 0.031961 |
| fast service | 0.036787 |
| hassle free | 0.041842 |
| app easy | 0.042814 |
| easy use | 0.046344 |
| helpful app | 0.056958 |
| friendly app | 0.065652 |

Customers had a **negative** experience when using the app for

1. The app was slow / not loading.
2. The app was not showing policy / claims details.
3. Customers were not able to add policy in the app.
4. Some were even not able to login.
5. The app was not working as per expectations.

Customers had a **positive** experience when using the app for

1. Hello doctor tele-consultation service was helpful
2. The app was easy / simple / friendly.
3. The app was fast / quick.

We performed separate modelling exercises for HDFC Ergo & Acko each. Important predictors of sentiments for both are the same as Industry level important features.

# 4. Business Insights

1. Companies need to investigate why some customers are facing slowness of apps. Is it because the app is heavy on memory or needs fast internet connection.
2. Customers faced issues in receiving OTP and some were not able add or link insurance policies in the app. Better integration with backend IT infrastructure can solve this issue.
3. Companies can increase the adaptation of mobile applications by making customers aware of the applications and benefits. This initiative was well received by customers as we have seen with ICICI Lombard.
4. Value added services like tele consultation with doctors can definitely boost the positive experience of customers in case of health insurance.