Predict the Next Big Thing

NLP Driven Forecasting System

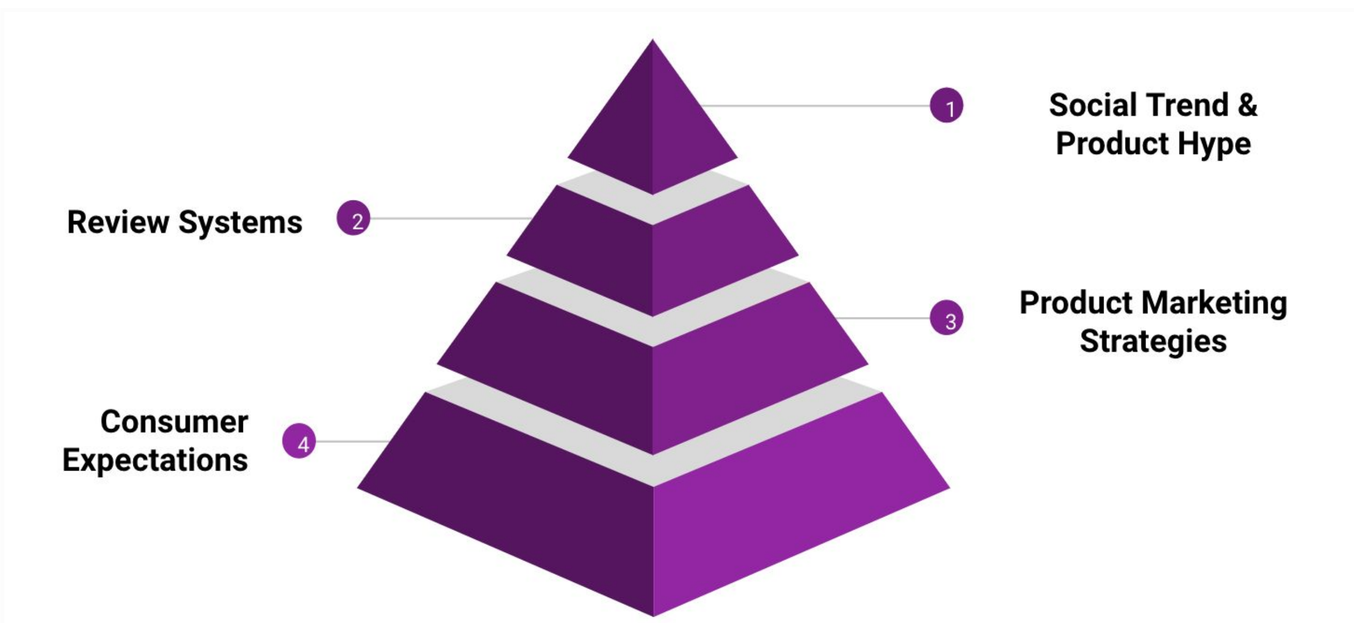
Team: Challenger

- Pranav Pawar
- Dipashree Patil
- Durga Sandeep

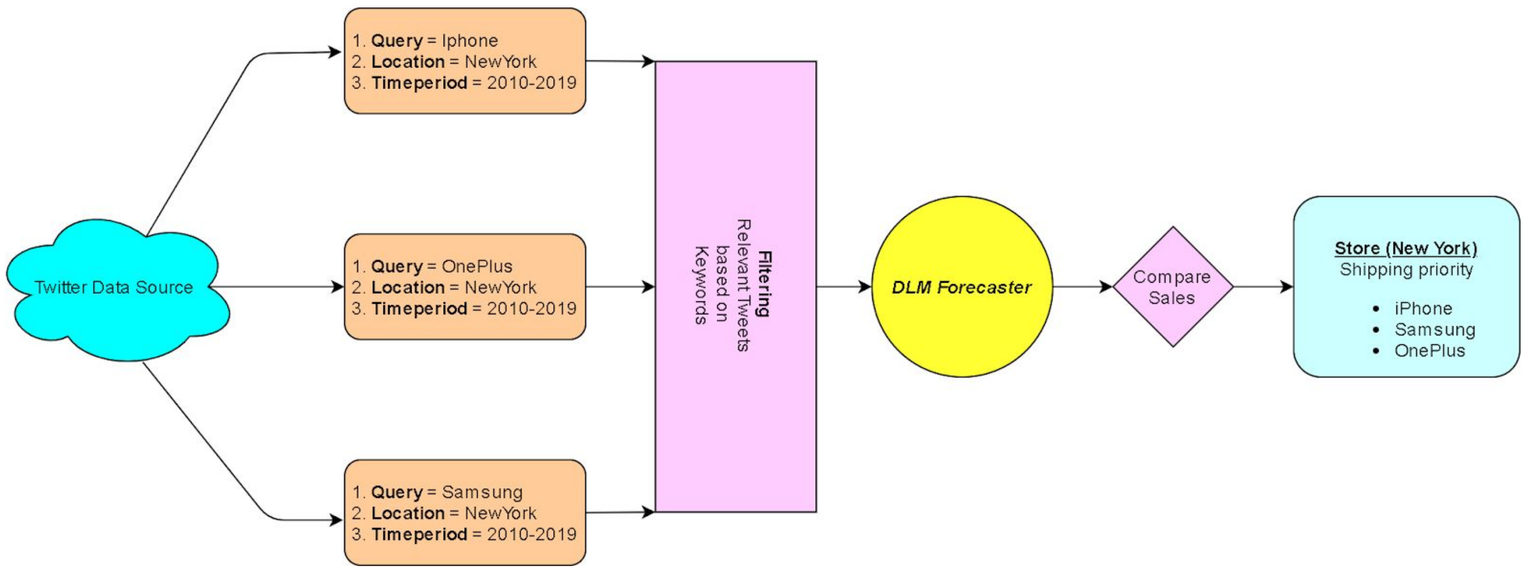
Overview -

- According to the Forbes report of 2014, Walmart executives reported they were leaving almost $3 billion on the table as a result of out-of-stocks.
- Though Walmart had taken the necessary steps to solve the problem, the problem still persists at the supplier level of the Supply Chain.
- We Propose an NLP driven Deep Learning system to forecast customer demand with precise accuracy.
- The system accounts for social trends with Twitter, and customer base behavioral analysis.

Responsible Factors



NLP Forecaster Pipeline



Identifying Social Trends (around a given location)

- We used "**Where On Earth Identifiers (WOEIDs)**". The **WOEID** database has been put together and made freely available by Yahoo!, and is used for geographic-based trends by Twitter.
- An example to identify the trends in the US.

```
# The Yahoo! Where On Earth ID for the entire world is 1.
# See https://dev.twitter.com/docs/api/1.1/get/trends/place and
# http://developer.yahoo.com/geo/geoplanet/

WORLD_WOE_ID = 1
US_WOE_ID = 23424977

# Prefix ID with the underscore for query string parameterization.
# Without the underscore, the twitter package appends the ID value
# to the URL itself as a special case keyword argument.

world_trends = twitter_api.trends.place(_id=WORLD_WOE_ID)
us_trends = twitter_api.trends.place(_id=US_WOE_ID)
```

- Review systems: Analyzing customer reviews from official twitter accounts. Top mobile reviewers like Marques Brownlee, Austin Evans.TysiPhoneHelp.
- Pandemic Effect: To manage the inventory of essential goods instead of electronics which has lesser demand during those periods.
- Promotional markdown events:
    - Samsung IPL, Oneplus music concert, Super Bowl, Labor Day Savings, Thanksgiving, Christmas, Newyear. These events have a huge impact on sales. So these can be manually or automatically fed into our pipeline(query search)

Filtering Tweets

Why filtering before "DLM Forecaster"?

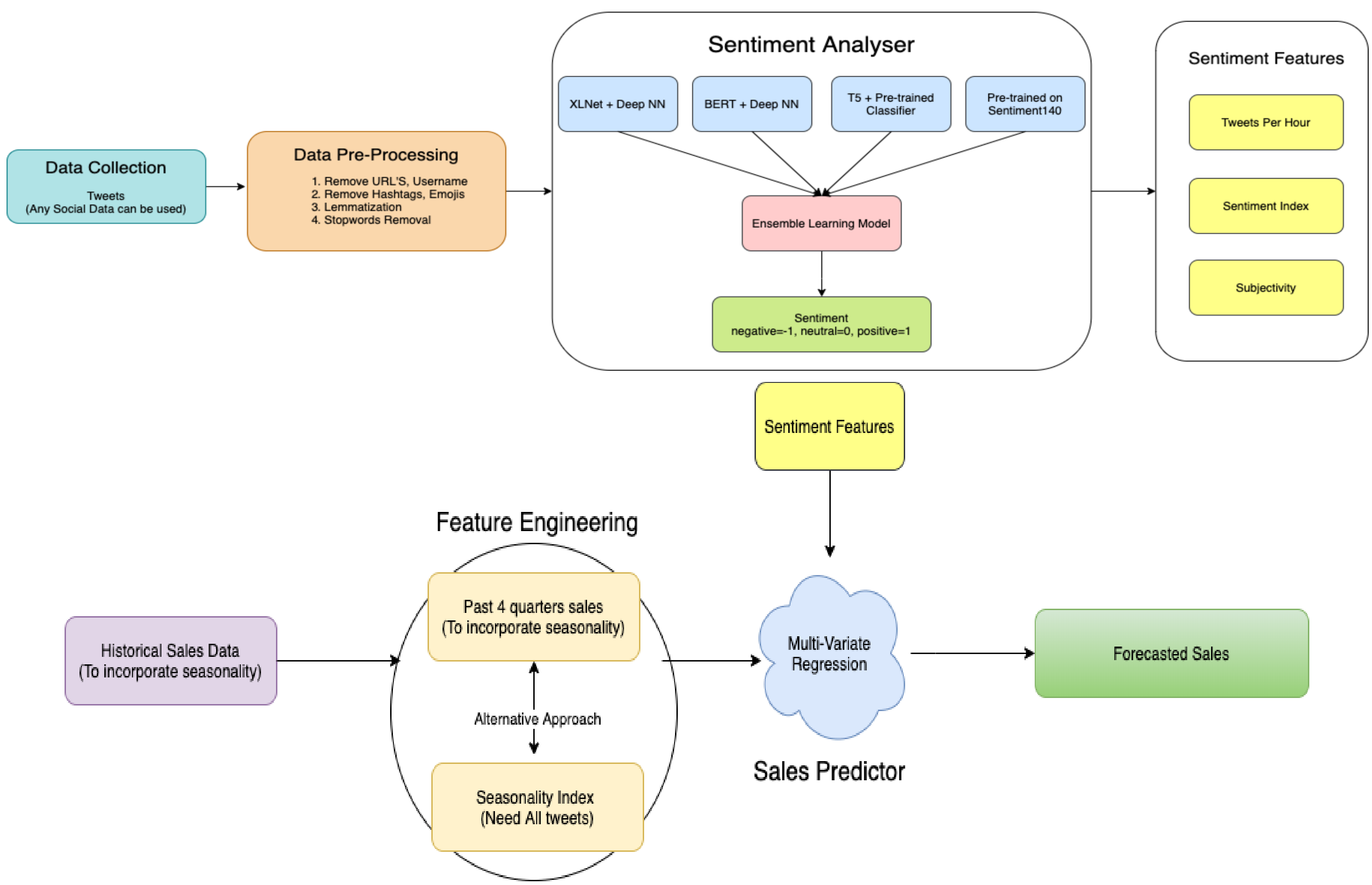
- For queries like "iPhone" and "OnePlus", they are unique names of their product which certainly implies Smartphones but queries like "Samsung" may imply not just mobiles but also other electronic products.

- Not all tweets are about the review of the smartphone, few tweets like posting images that are captured from the iPhone. These tweets will not be useful in forecasting. So, the intent of the tweets has to be taken care before feeding to "***DLM Forecaster***"

- Remember after getting social trends from twitter API, we extract the tweets with query = event + "iPhone" or "Samsung" or "OnePlus" and add them to the corresponding quarter's data which are extracted just using query = "iPhone" or "Samsung" or "OnePlus".

- **Observation:** After filtering out the tweets based on the intent, prediction error has been greatly reduced(nearly 10%) when compared to "no filter".

Deep Learning Mobile Forecaster



Sentiment Features Extraction

1. **Tweets-rate: Number of tweets referring to a particular product per day**

   Tweet-rate(product) = No. of tweets(product) / time (in days)

2. **Overall Sentiment**

   Sentiment = No. of positive comments / (No. of positive + No. of negative) comments

3. **Overall Subjectivity**

   Subjectivity = (No. of positive + No. of negative) comments / No. of neutral comments

4. **Polarity: PN ratio**

   PN ratio = No. of positive comments / No. of negative comments

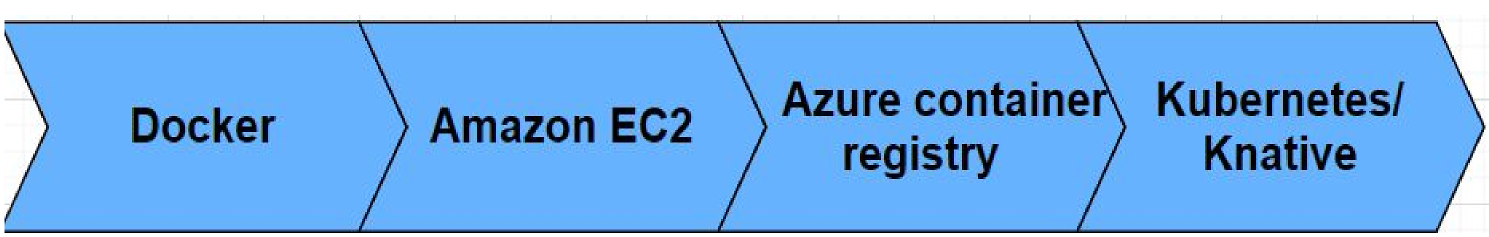
Deployment Platforms

1. To build an agile, secure & portable system Docker can be used to encapsulate and render the application along with Kubernetes to deploy and scale up the proposed system.

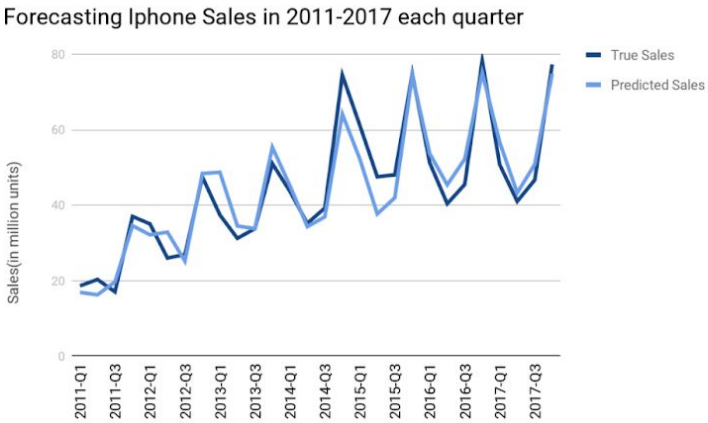
   Advantages: Docker is compatible with various PaaS(Platform as a Service) platforms like Heroku.

   Provides lightweight, fast, scalable, consistent & predictable systems.

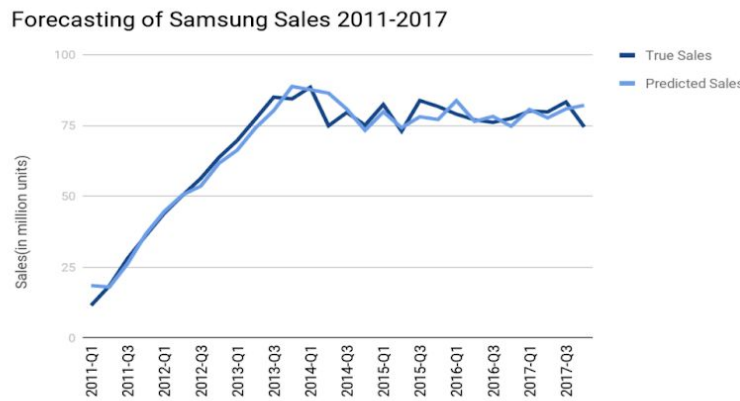
2. Alternative approach: Serving the model using REST(Representational State Transfer) APIs via Flask lightweight micro-framework on third party pipeline architecture allowing us to provide services quickly in a secure manner and dynamic update facility.



Training Phase Results

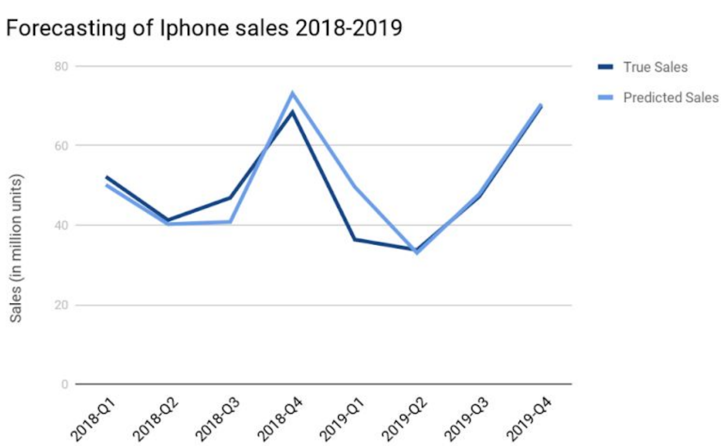


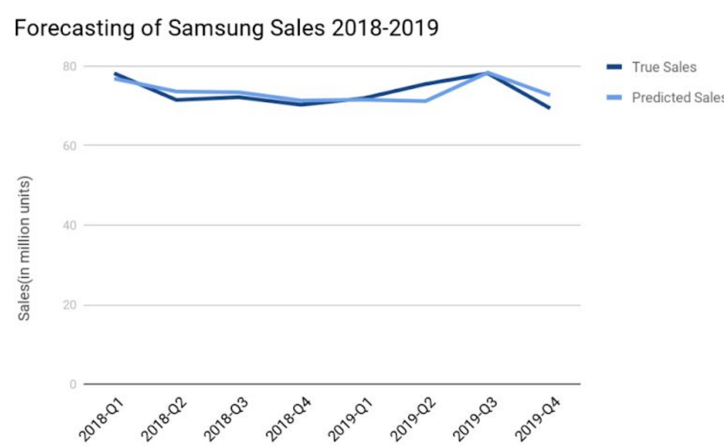
RMSE - 2.2565
MSE - 5.0919
MAE - 4.0932

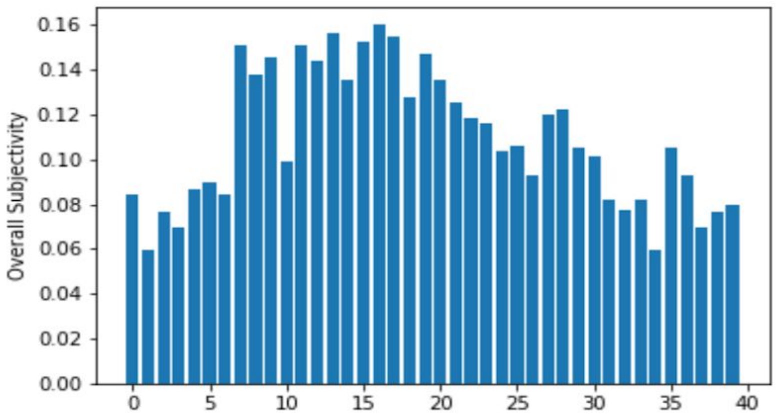


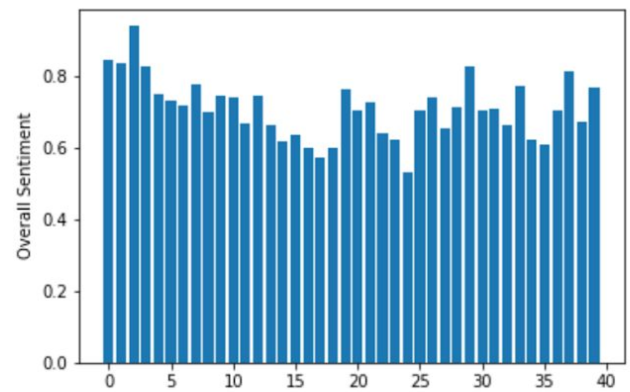
RMSE - 1.9813
MSE - 3.9258
MAE - 2.9778

Validation Phase Results





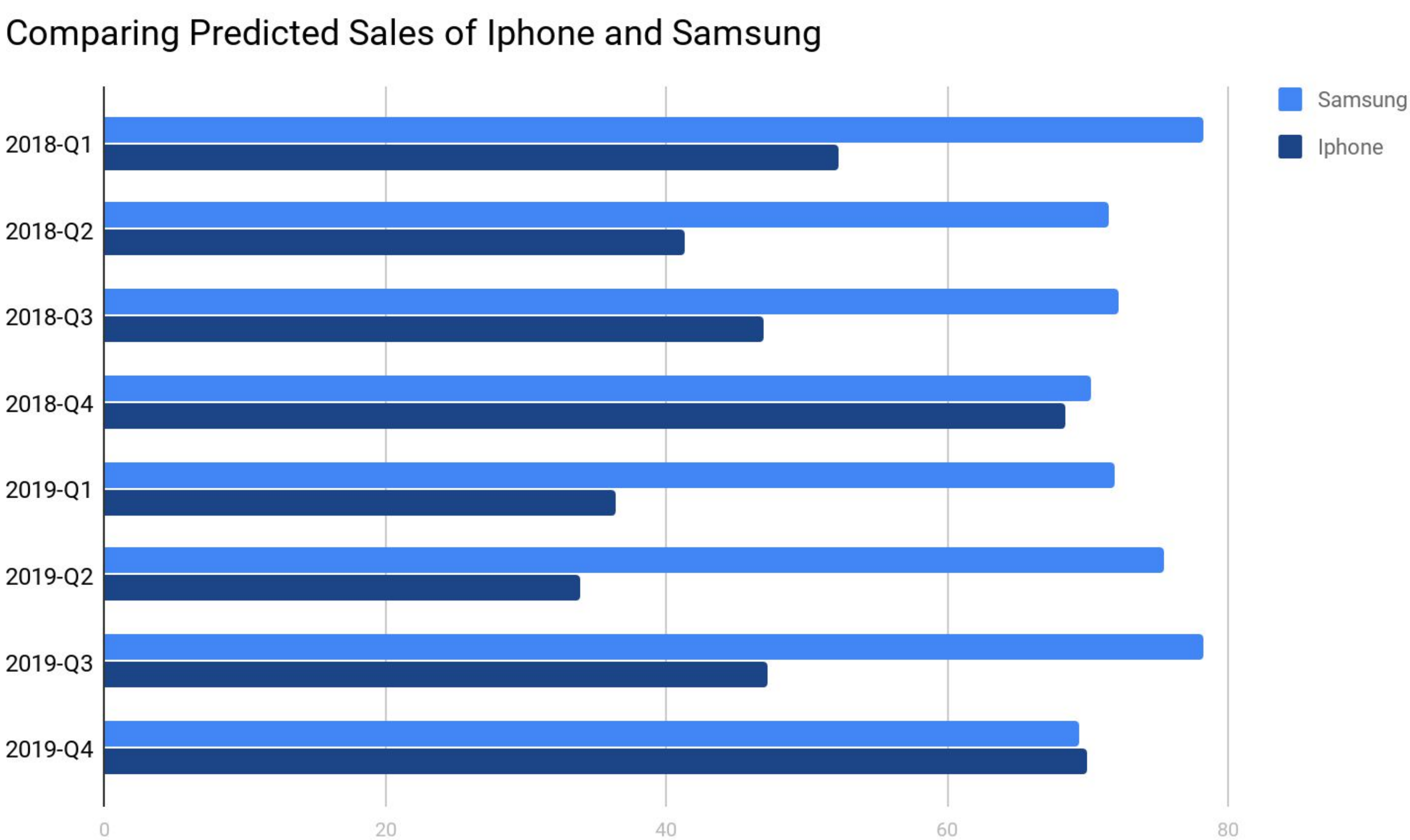




<u>Inference:</u> Above Bar plot corresponding to “Iphone”. Similar analysis for “Samsung” also.

1. Left plot shows a declining tendency over time suggesting that people are not as opinionated about iPhones as they used to be.
2. Right plot shows a declining tendency that indicates that people are still positive about iPhones but with the overall tendency is decreasing over time.
3. This is consistent with the fact that the latest versions of iPhone have not gained any major technological innovations but has shifted from "better" to "more" as in more CPU power, pixel density.

* Note: Subjectivity has very strong correlation with the Sales than Sentiment.



Further Improvements

1. OnePlus True Sales data is not public to access, so we weren't able to perform analysis on OnePlus.
2. Regional sales data of each outlet to perform frequency prioritization of sales and demand for each particular outlet.
3. Sales forecast & stocking up strategies to minimize the loss

References

- Asur S and Huberman B.A. "Predicting the future with social media", (IEEE, 2010, edn.)(Paper on predicting Movie Box office Sales using Social Media Data).
- Daniele. O'Leary "Twitter Mining for Discovery, Prediction, AND Causality: APPLICATIONS AND METHODOLOGIES"
- Zhao Jiang Lin and et al..“Learning to Learn Sales Prediction with Social Media Sentiment".
- VD Reijden P and Koppius O.R "The Value of Online Product Buzz in Sales Forecasting".
- Mining the Social Web, 2nd Edition by Matthew A. Russell
- Romero D, Galuba W, Asur S and Huberman B "Influence and passivity in social media", Machine Learning and Knowledge Discovery in Databases, 2011.
- BERT Implementation in PyTorch
- XLNet Implementation in Pytorch
- Text-to-Text Transfer Transformer" (T5) implementation using Transformers
- Sentiment140(1.6 Million tweets for Sentiment Analysis)
- Quarterly Sales data collection of Samsung and iPhone from "Statista"