Cost Estimation of Spam Call detection.

Introduction

Spam calls are now a daily part of our life. Many times, we choose not to answer unknown number because we are so sure of it being of an imposter claiming to be Nigerian prince and offering us a part of his treasure, which is hilarious because Nigeria is a republic and has no prince.

Hiya, provider of a spam monitoring service app, based on an analysis of its one-month data, found that each of its app users reported an average of 10 unwanted robocalls [1]. Another report published in the Washington Post said Americans got 26.3 billion robocalls in the year 2018 [2].

Scam calls are mostly done to steal money. One of every 10 US adults was victimized by a phone scam in the year 2018, with an average loss of $357 per victim [3]. Almost all of the mobile carriers are now coming up with some or the other measure for spam protection. One such approach uses Machine Learning to prevent malicious calls [4]. This is done by collecting a large-scale call-log database and then model an efficient machine learning algorithm to identify scam calls.

In this study, a cost estimate for the application of this technique in telecom industry is done. The cost estimate includes cost for all the necessary processes.

Ground Rules and Assumptions

- The cost estimate for app is only for the iOS platform.
- The estimate includes operation and maintenance cost for one year only.
- Data collection time is scheduled to be fixed to three months, based on analogy with [4].
- G&A, Project Management and Systems Engineering cost for the project has been estimated as the percentage of the total project cost.
- Cost for **promotion** element is not included in this cost estimate.
- The cost estimated is valid for USA only, the average values used are for the USA only.

Work Breakdown Structure (Context)

To establish a machine learning approach to detect scam calls, the most critical task is the collection of data. The collection of data is done by developing an iOS app that allows users to tag spam calls. The data is stored in a database after cleaning and pre-processing. Based on this data, Machine Learning Engineers develop appropriate ML-models which has the ability to predict spam calls. This is integrated with the app via an app update.

Work Breakdown Structure

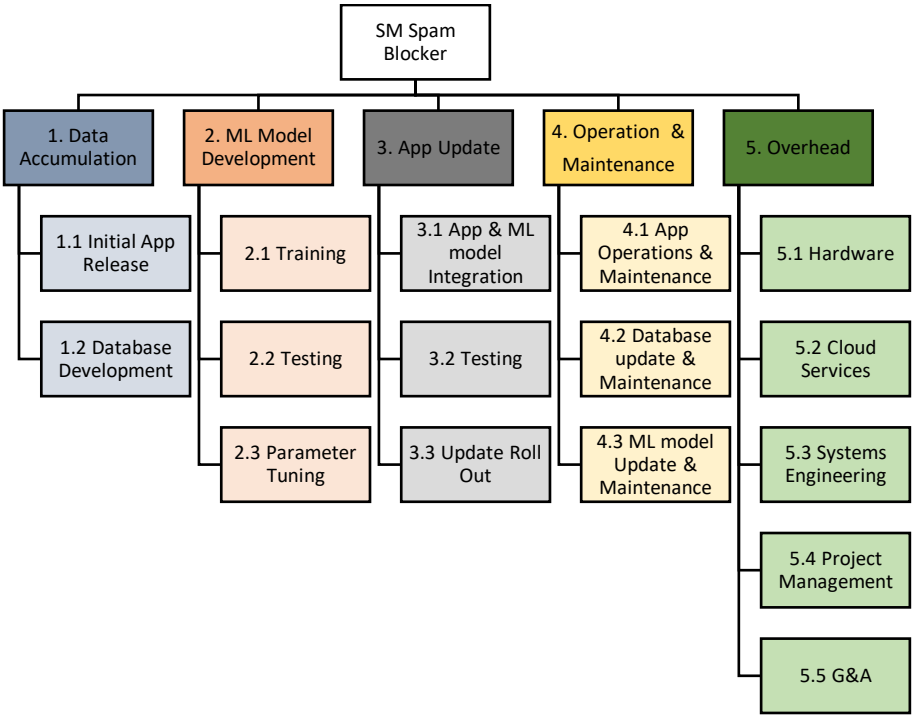


Figure 1: WBS - 1

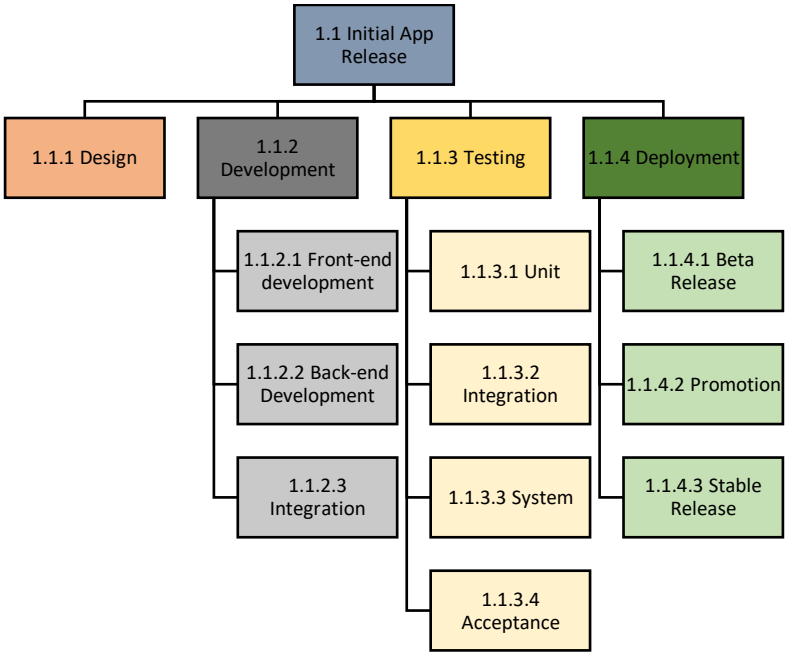


Figure 2:WBS-2

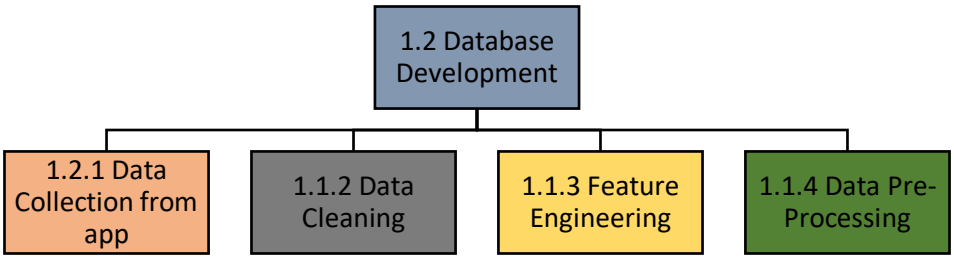


Figure 3: WBS-3

Note: The WBS is in disintegrated for better visibility.

The work breakdown structure is divided into five parts at the first level namely:

1. Data Accumulation

2. ML Model Development

3. App Update

4. Maintenance and Support

5. Overhead

Data accumulation consists of two phases, one is the deployment of the app that allows users to tag calls as spam. The other phase includes the organization, cleaning and pre-processing of the acquired data.

Based on this data, an appropriate model is developed. The development of the ML model consists of training, testing and parameter tuning steps. Once the ML-model is developed, an update is sent along to the app integrating ML framework.

The other two leaves in the first level are maintenance and overhead. The app, database and the model, all require maintenance and updates at regular interval of times. Also, there may be an unscheduled runtime issue that must be solved by the maintenance and support team.

The Overhead cost includes cost for hardware, cloud infrastructure, systems engineering, project management and G&A (General and administrative cost).

WBS Dictionary

WBS Element	WBS Element Definition
Front-end Development	Development of Graphical interface for the application.
Back-end Development	Development of data-access layer for the application.
Integration	Integrating of front and back end part of the application.

Unit Testing Integration Testing	Test to check the functioning of individual components of the software. Test to verify the interfaces between components against a software design.
System Testing	System testing tests a completely integrated system to verify that the system meets its requirements.
Acceptance Testing	Test to check operational readiness of the application.
Beta Release	Release of the beta version of the app, released to limited users to acquire feedback on the application.
Promotion	Promotion of the application, constitutes ads on the app store and other promotion strategy.
Stable Release	Release of the final version of the application.
Training	Training the machine learning model on training dataset.
Testing	Testing performance of the models on testing dataset.
Parameter Tuning	Tune hyper parameters of the models for better accuracy.
Hardware	Hardware requirements for the project, mainly includes computer systems required.
Cloud Services	Services made available through cloud computing provider's server.
G&A	General and Administrative costs.
Data Cleaning	Removing redundant, incomplete data points from the data-set.
Data Pre-processing	Processing data to the required format so as to fed to the ML pipeline.
Feature Engineering	Create additional features based of existing features.

Table 1: WBS Dictionary

Cost Summary

Top-Down Estimate

1. App v1.0 design, development and testing cost = $ 300,000 [5] [6]
2. 2. Disbursing information about the new app to existing telecom user (considering cost for iOS users only) = $200,000 [7]
3. Database development cost (Considering database development time for 3 months and a team of 10 data analysts) = $200,000 [8]
4. ML model development cost (Considering ML model development time of 3 months and a team of 5 Data Scientists) = $150,000 [9]
5. Cost to update app along with integrated ML model = $50,000
6. Overhead costs = +100% of direct costs
7. Maintenance & Support = +25% of development costs per year [10]

Overall cost* = $2,250,000 = 2.25 Million Dollars

*Overall cost is based upon the average cost and industry standards. The maintenance and overhead cost for only one year is included in calculation.

Bottoms-up Cost Estimate

Cost for the first element of the WBS (1. Data Accumulation)

The following section describes the cost distribution for element 1 of the WBS. The element 1 is divided into two parts, 1.1 Initial App Release and 1.2 Database Development.

Cost for element 1.1 of the WBS (Initial App Release)

The 1.1 element is further divided into four elements listed in the table below along with the median of high and low hours required for each of them. The values are based upon the App Development survey conducted by Clucth.co [11], which surveyed representatives from 12 leading mobile application development companies to determine cost ranges of building an iPhone app.

	Median Low (Hours)	Median High (Hours)
1.1.1 Design	52.125	238.5
1.1.2 Development	163.5	733.5
1.1.3 Testing	30	160
1.1.4 Deployment	5.125	11
Total Hours	252.75	1143

Table 2: Hours required to complete elements under element 1.1 of WBS [11]

The total time required for Initial app Release (element 1.1) is the sum of the elements 1.1.1 to 1.1.4. The above table shows the median low and high hours typically required for an app release (Total Hours row). The time for element 1, that is for Initial App Release is calculated by taking out the average of the median low hours and median high hours. This comes out to be around 700 hours.

The cost for the element 1.1 Initial App release is then calculated by multiplying the total number of hours with the average salary of the iOS developer in USA. The average salary for an iOS developer is summarized in the table below:

Source	Minimum Salary (annual)	Average Salary (annual)	Maximum Salary (annual)
Glassdoor	16000	102000	113000
Zip Recruiter	24000	84000	133000
Indeed	42000	124000	140000
PayScale	58000	84000	125000
Average	**35000**	**98500**	**127750**

Table 3: Pay for iOS developer in USA

The average of average salaries from all the sources is used for calculating the cost for the Initial App Release (element 1.1). The calculation is summarized below:

Average Hours for Initial Application development = **700 hours**

Average salary for iOS developers = $ 98500/ year = $98500 / 1920 hours = **$ 51.30 / hour**, assuming 5-day work week, 8 hours duration per day and 15 days of other leaves.

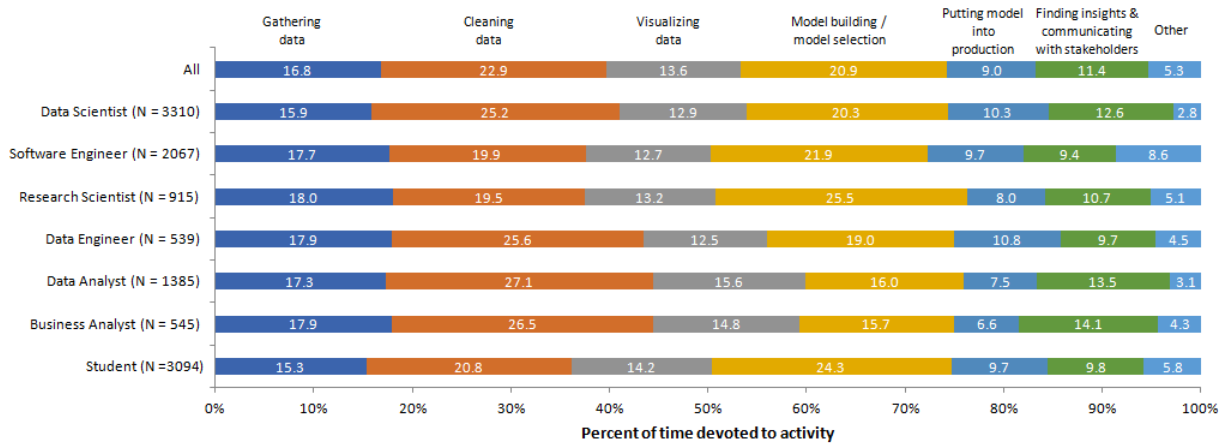
Average cost for 1.1 Initial App Release = $ 51.30 / hour * 700 hours = **$ 35, 910**

Cost for element 1.2 of the WBS (Database Development)

Approach and Source

This section describes the cost estimation of the element 1.2 (Database Development) of the WBS. The cost is again calculated by multiplying the time required for element 1.2 with the average salary of the professional who is going to undertake it, Machine Learning Engineer for this case. The same approach would be used throughout the report.

The time required for many of the elements (and this element) is calculated using the source [12] and approach mentioned below:



Note: Data are from the 2018 Kaggle ML and Data Science Survey. You can learn more about the study here: http://www.kaggle.com/kaggle/kaggle-survey-2018.
A total of 23859 respondents completed the survey; the percentages in the graph are based on a total of 15937 respondents who provided an answer to this question. Only selected job titles are presented.

Figure 4: Kaggle Survey

The time required for data accumulation is taken to be three months **(720 hours)**, based on the analogy with the similar project where machine learning was used to classify the benign and spam calls in China [4]. The time required for other parameters can be calculated based on the Kaggle survey [12], where around 24 thousand professionals have weighted (as percentage of the total project time) the time devoted to specific machine learning tasks. The weight for the same is summarized below:

Tasks	Weights
Data collection	0.168
Data Cleaning	0.229

Data Preprocessing and Feature Engineering	0.136
Model development	0.209
ML model and app integration	0.09
Others	0.168

Table 4: Weight Table

As the time required for the data collection task is fixed at 3 months, which accounts for 16.8 percent of the total time spent by the professionals, the time for other activites (or elements) can be calculated by unitary method. Since most of the other task are a function of the data collected, this unitary method is a good approximation for the time spent on each of the task.

Note: Same approach would be used for hours calculation for the other tasks as well. Will be referred as the unitary approach in this document from this point on.

Cost Calculation for element 1.2 of the WBS

The data used for training ML model to identify malicious calls in China, collected through similar tagging by users (using TouchPal application) methods in a time period of 3 months [4] is summarized in the table below.

Benign Calls	8929 million
Malicious Calls	73 million
Total Calls	9002 million

For this project, it is intended to gather data for the same time period as in the reference above. The data collected in this period can be less than the reference example. But, as the number of smart-phone users in USA is approximately one third of the number of smartphone users in China [4], it is assumed that the collected data would be able to capture all the necessary variations important for classifying spam and benign calls in USA.

Time devoted to data collection is 3 months ((3month*30 days*8 hours) 720 hours). The table below shows the time required for data cleaning, feature engineering and data preprocessing (tasks listed under 1.2) calculated based on the weights in Table 4

	Task	Weight*	Calculation (Hours)	Hours
1.2.1	Data Collection	0.168	0.168*(720/0.168)	720 hours
1.2.2	Data Cleaning	0.229	0.229*(720/0.168)	981 hours

1.2.3	Feature Engineering	0.136	0.136*(720/0.168)	583 hours
1.2.4	Data Pre-processing			

Note: Feature Engineering and Data Pre-processing is given as the visualization in the Kaggle Survey. Hence both are calculated jointly.

Below Table provides the average salary of machine learning engineer, the professional who is going to perform the (element 1.2) Database development task.

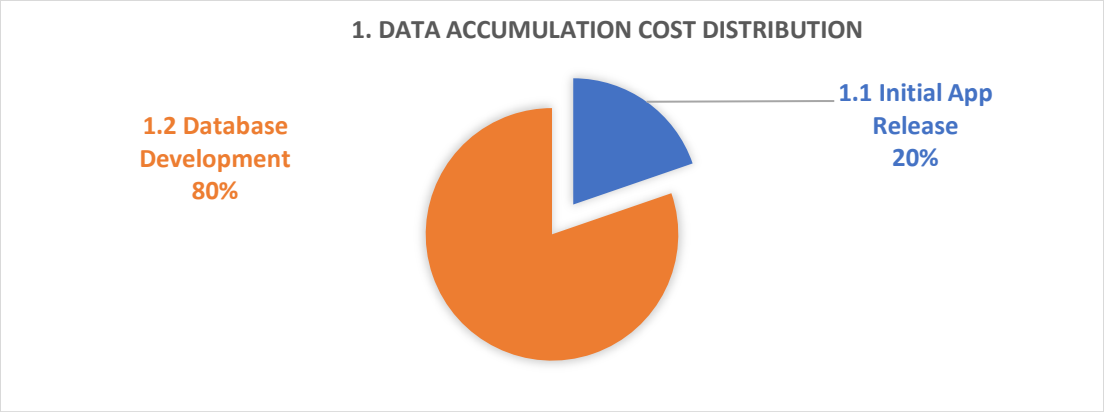
Source	Minimum Salary (annual)	Average Salary (annual)	Maximum Salary (annual)
Glassdoor	87000	121000	158000
Indeed	51000	135000	259000
PayScale	78000	112000	161000
Average	**72000**	**123000**	**193000**

Average salary for Machine Learning Engineer = $ 123000 / year = $ 123000 / 1920 hours = **$64 / hour**, assuming 5-day work week, 8 hours work duration per day and 15 days of other leaves.

The below table calculates the cost for each of the elements under element 1.2 (Database Development).

	Task	Hours	Calculation	Amount in $
1.2.1	Data Collection	720	720*64	46080
1.2.2	Data Cleaning	981	981*64	62784
1.2.3	Feature Engineering	583	583*64	37312
1.2.4	Data Pre-processing			
Total		2284		146176

The total cost required for the element 1, that is Data Accumulation is the sum of the cost of element 1.1 (Initial App Release) and 1.2 (Database Development). The same is summarized in the table and pie chart below.



1.1 Initial App Release	$35,910
1.2 Database Development	$146,176
Total	**$182,086**

Eighty percent of the Data Accumulation (element 1) cost can be traced back to database development.

Cost for the second element of the WBS (2. ML Model Development)

This section describes the cost estimated for the second element (ML model development) of the work breakdown structure. The second element cost is not divided further down the work break down structure as I was not able to separately collect the data for the time required for training, testing and parameter-tuning. It varies with project size and complexity.

For this project, time required for model development is calculated using the Kaggle survey data. The Kaggle survey reveals an average ratio of **0.804** (0.168/0.209) between the time required for data collection and the time required for the model development.

The time calculated for the data collection is 720 hours. Using the above ratio, the time required for model development of this project = (720/0.804) = 895.68 hours ~ **900 hours.**

Cost for the ML model development, considering model development time of 900 hours and $64 as the hourly rate of the Machine Learning Engineer = ($ 64 /hour) * (900) = **$ 57,600.**

	Task	Weight	Calculation (hours)	Time (hours)	Calculation (cost)	Cost (in USD)
2	ML model development	0.209	0.209*(720/.168)	900	900*64	57600

Cost for the third element of the WBS (3. Application Update)

This section of the report provides cost estimate for the third element of the WBS. The cost is estimated separately for the three elements under element 3 (Application Update). The three elements under Application update element are:

3.1 App and ML model Integration

3.2 Testing

3.3 Update Roll Out

Cost estimate for 3.1 App and ML model Integration.

The time required for the application and ML model integration is calculated using the weight derived using Kaggle survey data. The calculation is summarized in the table below.

The Kaggle survey reveals an average ratio of **1.867** (0.168/0.09) between the time required for data collection and the time required for application and model integration (putting model into production).

The time calculated for the database development is 720 hours. Using the above ratio, the time required for app and ML model integration for this project = (720/1.867) = **385 hours**

Cost for the app and ML model integration, considering time required be equal to 385 hours and $64 as the hourly rate of the Machine Learning Engineer = ($ 64 /hour) * (385) = **$ 24,640**

	Task	Weight	Calculation	Time (hours)	Calculation	Cost (in USD)
3.1	App and ML model integration	0.09	0.09*(720/.168)	385	385*64	24640

Cost estimate for 3.2 Testing and Update Roll Out.

The time required for testing and update roll out after the ML model and app integration is calculated using the Clutch Survey Data, mentioned before.

The median high and median low hours required for the testing and update roll out is taken to be the same as the time taken for application testing and application deployment. The survey results for the selected part are summarized in the table below.

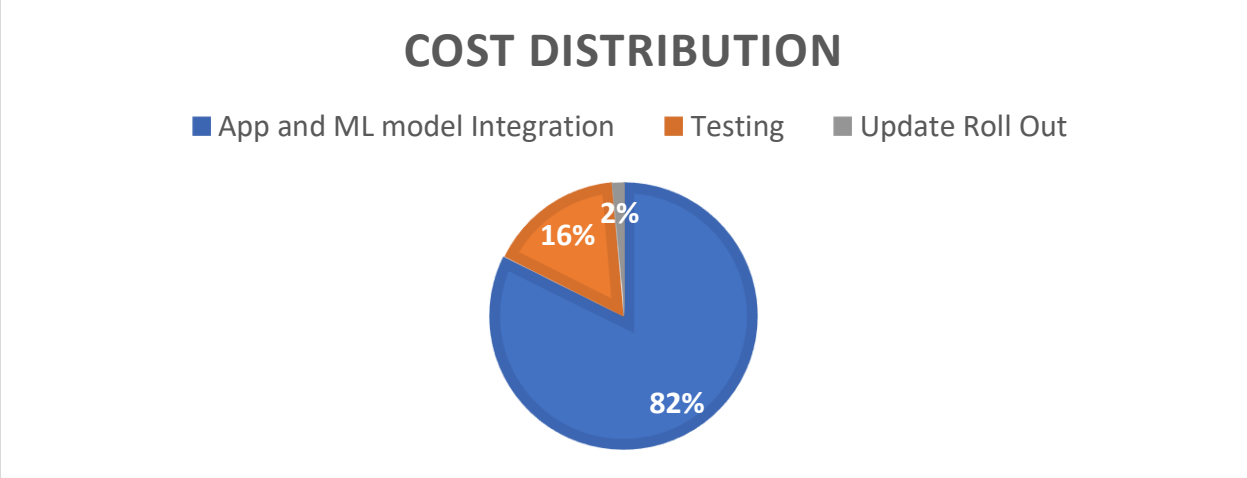
	Median Low (Hours)	Median High (Hours)	Average (Hours)
3.2 Testing	30	160	95
3.3 Update Roll Out	5.125	11	8.0625

The cost for testing and update roll out is calculated by multiplying the hours required for the project with the hourly rate of the iOS developers ($ **51.30 per hour)**. The calculation for the same is summarized in the table below.

Task	Average hours	Calculation	Cost (USD)
3.2 Testing	95	95*51.3	4873.5
3.3 Update Roll Out	8.0625	8.0625*51.3	413.6

The total cost for the element 3 is summarized in the table and pie chart below. It is calculated by adding cost for 3.1, 3.2 and 3.3 elements of the WBS.

Task	Average hours	Calculation	Cost (USD)
3.1 App and ML model Integration	385	385*64	24640
3.2 Testing	95	95*51.3	4873.5
3.3 Update Roll Out	8.0625	8.0625*51.3	413.6
TOTAL			251687.1



<u>Cost for the fourth element of the WBS (3. Maintenance and Operations</u>)

This section estimates the cost of element 4 of the WBS, Maintenance and Operations. The cost is estimated separately for the three elements under 4 Maintenance and Operations. The three elements under Application update element are:

4.1 Application Operations and Maintenance
4.2 Database update and maintenance
4.3 ML model update and maintenance

Note: The cost for maintenance and operations is calculated for one year only.

The application operation and maintenance are assumed to be performed by the applications and maintenance engineer. The database update and maintenance are assumed to be performed by the database administrator and the ML model update and maintenance are assumed to be performed by the ML engineer. The pay rates for the application operation and maintenance engineer and database administrator is summarized in the tables below. The pay rates for the Machine Learning Engineer are summarized in the in the previous part of the report.

Source	Low (annual)	Mean (annual)	High (annual)
Glassdoor	$64,000	$89,000	$161,000
Indeed	$25,000	$78,000	$108,000
Payscale	$49,000	$69,000	$97,000
ZipRecruiter	$45,000	$90,000	$117,000
Average	**$45,750**	**$81,500**	**$120,750**

Source	Low (annual)	Mean (annual)	High (annual)
Glassdoor	$65,000	$87,000	$118,000
Indeed	$32,000	$93,000	$184,000
Payscale	$46,000	$73,000	$110,000
ZipRecruiter	$52,000	$101,810	$145,500
Average	**$48,750**	**$88,700**	**$139,250**

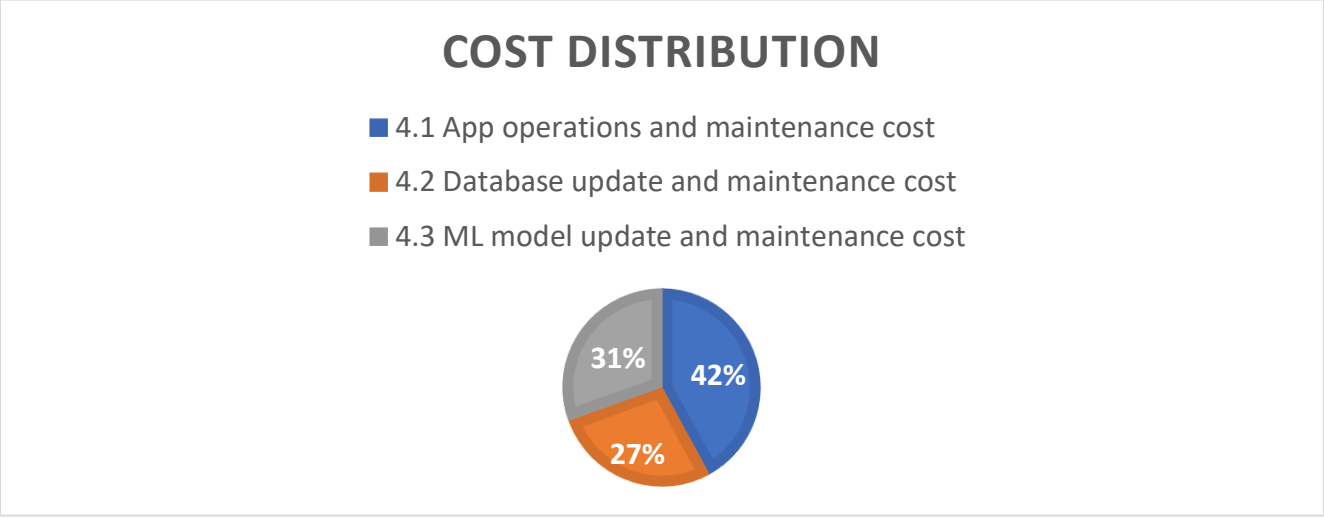
The structure is decided on the consultation with my advisor, Nilanjana Ghosh who has worked in the IT industry for three years and based on the assumption that [4] Maintenance and Operations would follow agile methodology. The structure can vary based on the complexity of the application. The typical number of people in a scrum team, compiled by various sources is summarized in the table below.

Source	People in Scrum team
M. Cohn, "Team Structure," in Succeeding with Agile: Software Development Using Scrum, Addison-Wesley Professional, 2009, pp. 177-199.	5-9
J. Sutherland, The scrum papers: Nuts, bolts, and origins of an agile process, Cambridge: scruminc.com, 2012.	7
Practical Scrum-Scrum Team: Way to Produce Successful and Quality Software, 2013 13th International Conference on Computational Science and Its Applications.	3-9

The structure for app maintenance and operations team along with the cost for each element under element 4 App maintenance and update is summarized in the table below.

	Team	Number of members	Average salary (annual)	Calculation (cost)	Total Cost
4.1	App operations and maintenance team	5	$ 81500	5* 81500	$ 407500
4.2	Database update and maintenance team	3	$ 88700	3* 88700	$ 266100
4.3	ML model update and maintenance team	3	$ 98500	3* 98500	$ 295500
TOTAL					$ 969100 / year

The pie chart below, shows the cost distribution for the element 4 of the Work Break down Structure. The cost is approximately homogenously distributed among.



Cost for the fifth element of the WBS (5. Overhead)

This section estimates the cost of element 5 of the WBS, Overhead. The cost is estimated separately for the five elements under 5 Overhead and is then added to get the estimate for Overhead. The five elements under Overhead element are:

5.1 Hardware
5.2 Cloud Services
5.3 Systems Engineering
5.4 Project Management
5.5 General and Administrative

5.1 Hardware Cost

The hardware cost is the cost of the laptops purchased for the project. The estimation for the laptops required for the project is capped at thirty. Hardware cost is calculated below. The cost for each laptop is extracted from vendor quotes [amazon, BestBuy].

Cost = 30 * $ 1300 = $ 39000 ~ **$ 40000**

5.2 Cloud Services Cost

The cloud requirement for the project is estimated to be 50TB. The rate for standard storage on Google Cloud is $0.026 per GB per month. Following the google cloud rate, the cloud services cost for 50TB storage for one year is calculated below.

Cost = (50 * 1024 GB) *($ 0.026/month * GB) *(12 month) = **$ 15975 / year**

5.3 Systems Engineering Cost

The systems Engineering cost for the project is taken as to be the 15 percent of the project cost [industry average]. The cost for the same is calculated below.

Cost = 0.15 * (251687.1 + 57600 + 182086 + 969100) = 0.15 * (1460473.1) = **$ 219070**

5.4 Project Management Cost

The project management cost for the project is taken to be the 8 percent of the project cost [ref: Byrne, J.P. (1999). Project management: how much is enough?]. Same is calculated below.

Cost = 0.08 * $ 1460473.1 = **$ 116838**

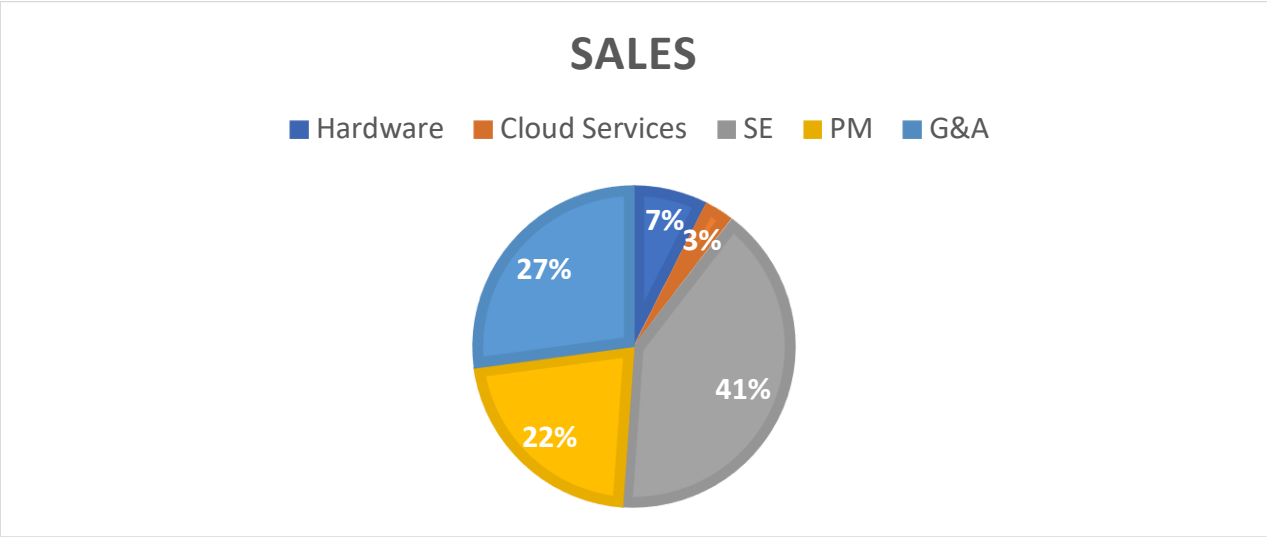
5.5 General and Administrative Cost

The G&A cost for the project is taken to be the 10 percent of the project cost [nationalservice.gov]. Same is calculated below.

Cost = 0.10 * $ 1460473.1 = **$ 146047.31**

The total overhead cost is calculated in the table below. The pie chart below summarizes the cost distribution for Overhead. The major cost drivers in Overhead cost are systems engineering, project management and G&A.

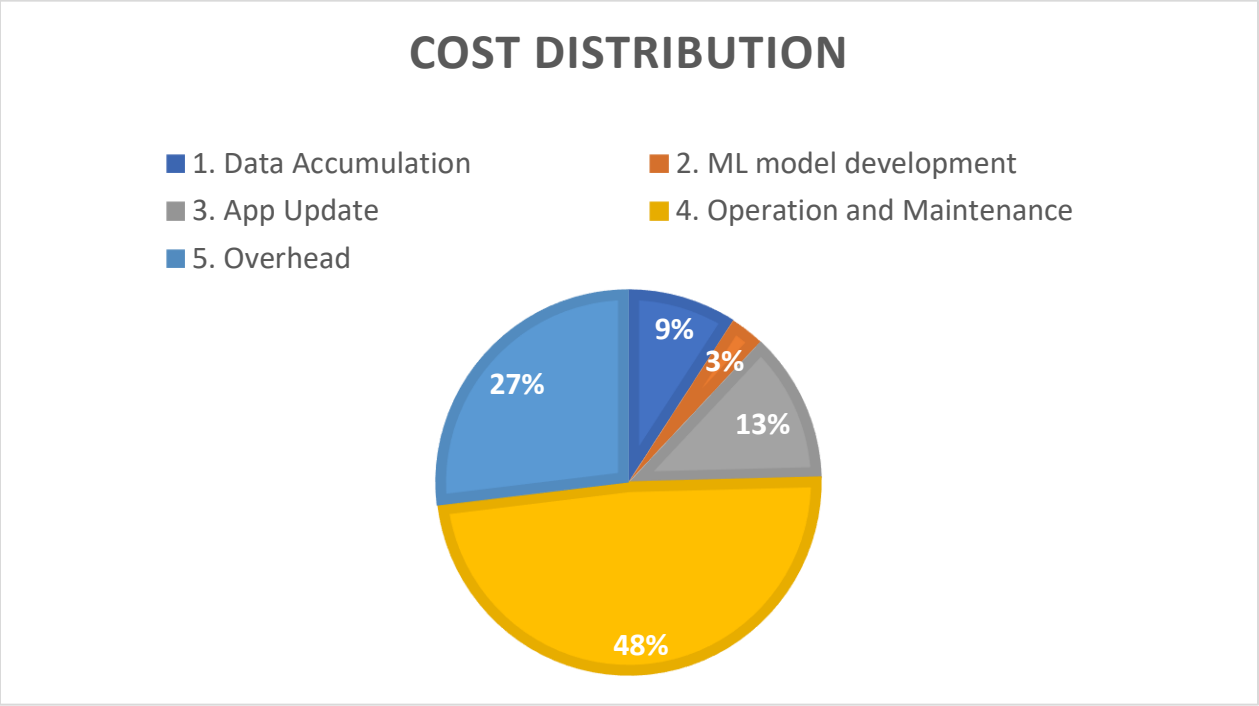
	Cost (USD)
5.1 Hardware	40000
5.2 Cloud Services	15975
5.3 Systems Engineering	219070
5.4 Project Management	116838
5.5 G&A	146047
Total	**537930**



Overall Project Cost

This section describes the estimated overall project cost which is calculated by adding up the estimated cost for the all the first level element of WBS. This is summarized in the table below.

	Cost (USD)
1. Data Accumulation	182086
2. ML model development	57600
3. App Update	251687
4. Operation and Maintenance	969100
5. Overhead	537930
TOTAL	**1998403**



The major cost driver for the project, as can be seen from the graph above is Operation and Maintenance. It accounts for almost half of the total cost. The cost estimate for the project, using bottoms up approach, is coming out to be $ 1998403, approximately equal to 2 million. The estimated through the top down approach was $ 2.25 million. The difference between the two can be due to the fact that the bottoms up approach didn't take the amount for promotion into account while the Top down approach did take a certain amount cost for the promotion element.

Top down Estimate presented in kick off presentation = $ 2.25 Million

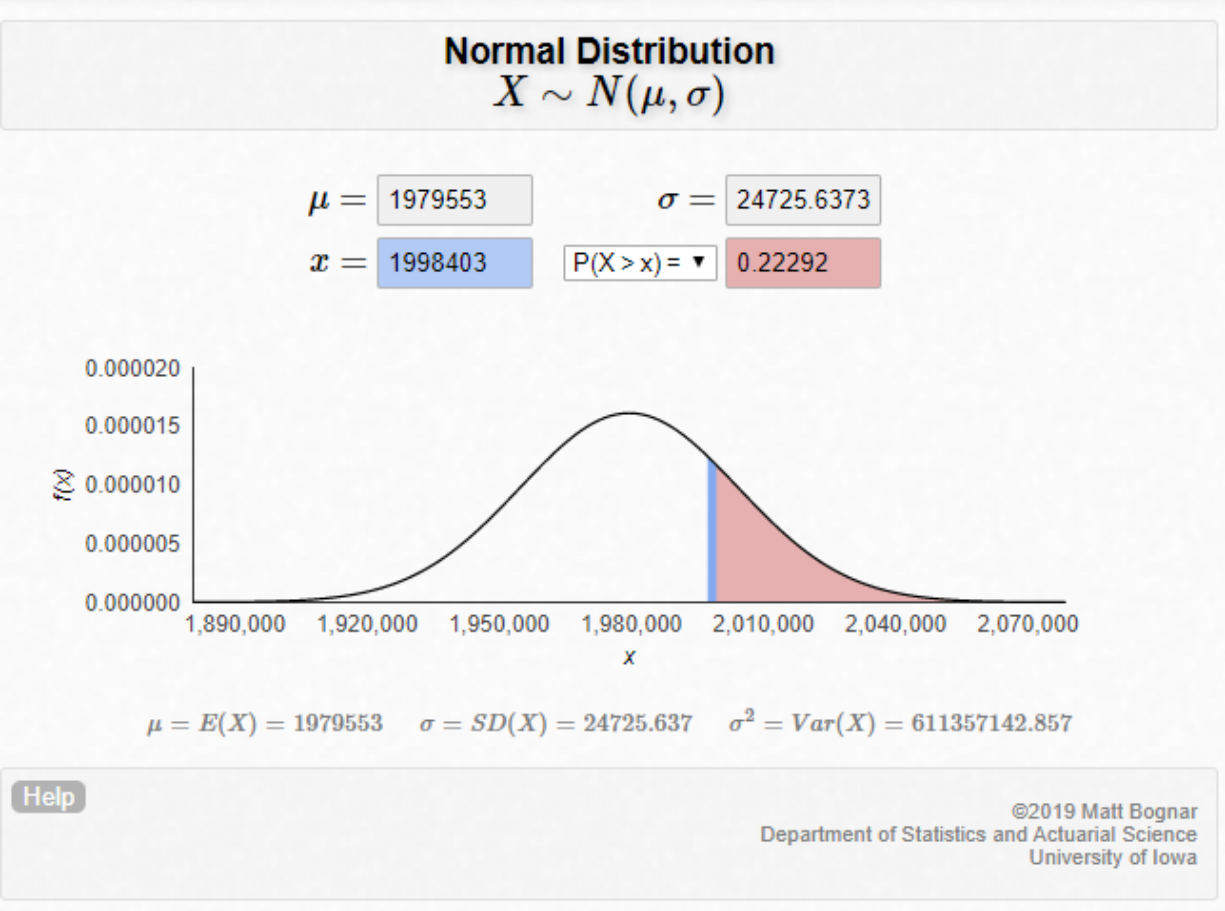
Bottoms up cost estimate = $ 1998403 ~ $ 2 Million

Risk Analysis

The risk analysis is conducted for the fourth element of the WBS, operations and maintenance as it is the largest cost driver. The risk analysis is done using central limit theorem.

The cost for the element four is varied by taking different salaries for the machine learning engineer, database administrator and the app operations and maintenance engineers. Cost of the remaining elements are kept same. Another approach could have been by varying the number of people in the team, I didn't take this into account. The risk analysis is limited in that sense.

The mean, standard deviation and the probability of cost overrun is summarized in the figure below. The risk of cost overrun for the estimated top down cost of **$1998403** is **22.3%.**



References

Note: Other reference are mentioned within the report.

[1] Based on report put forward by Hiya, compiled by Washington post.

[2] Report compiled by Washington Post (https://www.washingtonpost.com/technology/2019/01/29/report-americans-got-billion-robocalls-last-year-up-percent/?arc404=true&noredirect=on)

[3] Truecaller Blog (https://truecaller.blog/2018/04/26/truecaller-insights-usa-2018/)

[4] A machine Learning approach to prevent malicious calls over telephony networks, (https://arxiv.org/pdf/1804.02566.pdf)

[5] Figuring the costs of custom mobile business app development by Adriana Neagu, Formotus

[6] App development costs: The ultimate guide to app budget by app type by Ken Yarmosh, Savvy

[7] Fikusu's 2015 research

[8] Average Data Analyst Salaries in Washington DC area, Glassdoor

[9] Average Data Scientist Salaries in Washington DC area, Glassdoor

[10] Computer Economics, Metrics for IT Management

[11] Clutch.co, app development survey 2017

[12] Kaggle Survey