Implementation of Machine Learning for Procurement Assistance

A Project Presented to

The Faculty of the Department of Industrial and Systems Engineering

San José State University

In Partial Fulfillment of the Requirements

For the Degree Master of Science Engineering Management

By

Samarth Bhargava

SAN JOSÉ STATE UNIVERSITY

The Undersigned Project Committee Approves the Project Titled

Implementation of Machine Learning for Procurement Assistance

By

Samarth Bhargava

APPROVED FOR THE DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING

S. Ayca Erdogan, Department of Industrial and Systems Engineering

Contents

List of Figures/Tables

List of Tables

Abbreviations

SAP: Systems, Applications, Product in data processing.

ML: Machine Learning

ERP: Enterprise Resource Planning

HPE: Hewlett Packard Enterprise

SNL: SAP Labs Network

ELK: Elasticsearch, Logstash, Kibana.

ETL: Extract, Transform and Load

AWS: Amazon Web Services

EC2: Elastic Cloud Compute

CNN: Convolutional Neural Network

KPI: Key Performance Indicator

TP: True Positive

FP: False Positive

FN: False Negative

TN: True Negative

HPC's: High Performance Computing

VM: Virtual Machine

API: Application Program Interface

HTTP: Hypertext Transfer Protocol

DNS: Domain Name System

Abstract

In any organization, there is always a need to make the process more efficient and less expensive. The objective of the project was to implement machine learning for Procurement Assistant at SAP Labs LLC. The report starts with brief description of the company and the business problem being tackled. After that the reports dive into how analytics can be used to determine which model to use and what data to be used to train it. The report then talks about different feature extraction done on the data to come up with an appropriate model and fine tuning it later. The focus then is building the model and releasing it successfully to the production environment.

The project plan is defined, and the requirements are documented to build the product accordingly. The responsibilities of a Machine Learning Quality Analyst included to overlook over the GAP analysis, planning the sprints and making sure to achieve them. Performing statistical analysis and deciding on the model was a major role played by the Quality Analyst. Analysis of both short-term and long-term benefits for the company and the customers after the product was deployed and building visualization tools for the same was a major task that followed once the project was up and running.

The project was a success and was greatly accepted by the customers of SAP worldwide.

1. Introduction

To grow rapidly involving Machine Learning these days is a must for any organization. SAP Labs is no different from that, rather they are one of the best in their sector. For this amazing Machine Learning to happen a lot of analysis goes into work. Models are made using data gathered from the real world and are meant to predict future. Analyzing data in the right way is the key to great success and is accepted by all the companies. Machine Learning tries to automate tasks which before required human intervention, a way where we can teach machine to learn using data and then expects it to perform as a human would have or in some cases better than humans.

An organization's growth is only limited by the amount of resources it has. If the resources would have been unlimited there would have been no competition in the world. In an organization strategic purchaser is responsible for the procurement of materials for smooth functioning of the organization. Deciding which material to buy and which not to buy is an inexpensive and important task and needs to be done with utmost precision to sustain maximum profits out of the investments done by the organization for the materials. To solve this business problem a team of designers and engineers are dedicated at SAP Labs to come up with a Procurement Assistant implemented with Machine Learning to help organizations to reach their maximum potential.

Maximizing profits for SAP customers while minimizing cost and wasted materials during the procurement process was the goal of the project.

1.1 About SAP Labs.

SAP SE is a European German based global organization. System, Application and Products in Data Processing. It is a multinational software corporation which provide business solutions enterprise software to its customer to manage their business. Its headquarters is situated Walldorf, Baden-Württemberg, Germany along with regional offices around 180 countries having over 330,000 customers. In 1973, the main business item was propelled. It was called SAP R/98 and offered a typical framework for numerous undertakings. This allowed the utilization of a brought together information stockpiling, enhancing the support of information. From a specialized perspective, in this manner, a database was necessary.

In 1976, SAP GmbH was established, and moved its central station the next year to Walldorf, Germany. After three years, in 1979, SAP propelled SAP R/2, extending the capacities of the framework to different territories, for example, material administration and generation arranging. In 1981, SAP offered a re-structured item for sale to the public. Be that as it may, SAP R/2 did not enhance until the period somewhere in the range of 1985 and 1990. SAP discharged the new SAP R/3 of every 1992. SAP created and discharged a few adaptations of R/3 through 1995. By the mid-1990s, SAP pursued the pattern from centralized computer figuring to customer/server designs. The improvement of SAP's web procedure with mySAP.com upgraded the idea of business forms (mix through Internet). subsequently, R/3 was supplanted with the presentation of SAP ERP Central Component (ECC) 5.0 in 2004. Architectural changes were additionally made to help a venture benefit engineering to progress clients to an administrations arranged design. The most recent variant, SAP ERP 6.0, was discharged in 2006. SAP ERP 6.0 has from that point forward been refreshed through SAP upgrade packs, the latest: SAP improvement bundle 8 for SAP ERP 6.0 in 2016.

Since 2012, SAP has procured a few organizations that offer cloud-based items, with a few multibillion-dollar acquisitions seen by investigators as an endeavor to test contender Oracle. In 2014 SAP purchased Concur Technologies, a supplier of cloud-based travel and cost the executives programming, for $8.3 billion, SAP's most costly buy to that date. Analysts' responses to the buy were blended, with Thomas Becker of Commerzbank addressing whether Concur was the correct decision for SAP, while Credit Suisse considered the obtaining a "forceful" move.
In 2014, IBM and SAP started an association to offer cloud-based services. Likewise, in 2015, SAP additionally banded together with HPE to give secure crossover cloud-based administrations running the SAP platform. Both HPE and IBM give framework administrations to SAP, and SAP runs its SAP HANA cloud arrangement to finish everything. SAP has reported extra associations with Microsoft with the end goal to give clients apparatuses for information representation, and additionally enhanced portable applications.

SAP surpassed its income projections because of the development in its cloud business and the achievement of SAP HANA. The development can likewise be in part credited to the acquisitions of Concur and Fieldglass. The organization reported plans in 2016 to put intensely into innovation identifying with Internet of Things (IoT) as a major aspect of a procedure to benefit from the development in that showcase. For that reason, €2 billion is gotten ready for interest in pertinent parts before the finish of 2020. SAP will likewise dispatch another product offering called SAP IoT, which "will join a lot of information from things associated with the Internet with machine learning and SAP's ongoing database S/4 HANA.

SAP Labs are SAP's center R&D elements, creating and always enhancing key SAP arrangements. These innovative work areas are deliberately situated in cutting edge groups far and wide and mirror SAP's way of life of decent variety and advancement. Every Lab has a specific spotlight on explicit applications, innovation, or the market. With 20 Labs in 17 nations, SAP Labs Network (SLN) drives thought initiative comprehensively and in neighborhood biological communities, enabling SAP to develop, develop, and succeed.

SAP Labs are strategically located in all the high-tech clusters around the world. The four most conspicuous labs of SAP SE are situated in Germany, India, China and the US. Labs Walldorf was established in 1972 and turned into SAP's essential area. Toward the starting, the focal point of SAP's extension was entering very created IT markets; in 1993 Palo Alto turned into a piece of SAP Labs. Planning to procure skilled workers, SAP opened another lab in Bangalore in 2003.

1.2 Project Requirement

With the end goal to recognize the business issue and need, a GAP analysis was performed at SAP by talking to every customers and stakeholders – building supervisors, program chiefs, improvement groups (document frameworks, stage, framework the board, UI, biological communities, designing administrations), quality confirmation group, client bolster group, item supervisory group, executives, senior executives and building pioneers.

GAP analysis is a formal investigation of what a business is doing presently and where it needs to be later. It includes deciding, reporting, and enhancing the contrast between business prerequisites and current capacities.

Gap analysis consists of (1) identifying future desired state (2) analyzing the current state (3) highlighting the gap and defining an approach to bridge the gap. Below is the summary of the gap analysis results-

Desired State:

- Adding new features to deliver increased value to our customers.
- Implementing Machine Learning for the procurement assistant solution offered by SAP.
- Build a recommendation system to make smarter decisions.
- Reduce costs and waste for the customers.
- Contracts can be created on a regular basis, based on recommendation from the historical data.

Current State:

- Contract creation for procurement of material takes lots of time.
- The system is prone to human errors.
- Enough historical data available to create a Machine Learning model.
- Contract creation generally happens on an annual basis creating a backlog for the customers if procurement fails.

Required Capabilities:

- High quality, real time recommendation for the customers.
- Feedback mechanism to be installed for constant improvement.
- Flexibility to create contracts when required.
- Significant reduction in time taken to create a contract.

- A dashboard to track all the contracts.

Plan:

- Analyze data and identify required features.
- Clean the data identified.
- Develop dashboard to visualize data.
- Distribute tasks among different team.
- Develop machine learning model using the data identified.
- Integrate the model with the existing procurement assistant.
- Test and monitor the results.
- Develop a continuous learning and improving model.

After performing GAP analysis, the immediate goal was to design the product while meeting the specified deadlines and delivering a quality product using the data identified. The product should provide meaningful recommendations to the customer and should implement a continuous improvement mechanism.

2. Background

SAP’s clients range from small businesses to major multinational conglomerates and it provides them business solutions. One such solution they provide is the ability to track and maintain and negotiate the procurement of materials needed by the businesses. These businesses generally work out yearly contracts for all the raw materials they need at once and feed it into the system. There were multiple issues with this approach. Firstly, given it wasn’t automated and done by the strategic buyer of the company, it was error prone. Secondly, it resulted in competition and material exhaustion because multiple vendors were aware of when contracts expire. This also made the contracts expensive and resulted in multiple sour exchanges that impacted businesses.

The solution that was sought for this problem was to build a recommender that can analyze the historical material procurement data and decide when the best contracts are available to buy and when it is optimal to switch. The system used data from SAP's product catalog that contained 200,00 types of products with multiple sub products totaling over 10 million types of products. It also used data specific to each business to understand their material needs. Combined, the system can produce ideal recommendations to buy materials at the cheapest rate before at a fixed amount of time before a contract expires. It also suggests an ideal amount of time to use as the term for the contract based on historical trends in procurement data.

2.1 Project Timeline

Project timeline is demonstrated in the below Table 1.

Table 1. Project Timeline

Item	Task	Task Owner	Initial Estimate (Total Sprint Hours = 50 x 5) 250
Building Machine Learning Prediction Model	Building data visualization	****	30
	Gathering required data and feature extraction	****	30
	Selecting the right model	****	30
	Training and testing	****	30
	Model Tuning	****	30
Performance analysis of Built System	Gathering performance data	****	15
	Making an API	****	15
	Analysis	****	15
	Visual Information Gathering	****	15
	Introduction of feedback to the model	****	15
Future roadmap	Provide possible improvements	****	15
	User Recommendations	****	10
Total Available Hours During Sprint:	**250**		
	****- Manager at SAP		

3. Building Machine Learning Prediction Model in Python

The need to generate better prediction and forecasting models using the data available has led to various statistical and linear algebraic concepts being applied to computational algorithms. This new area is colloquially known as Machine Learning where systems try to learn about the features of the available data and automatically improves its accuracy.

A general concept of machine learning is demonstrated by the Figure 1 below:

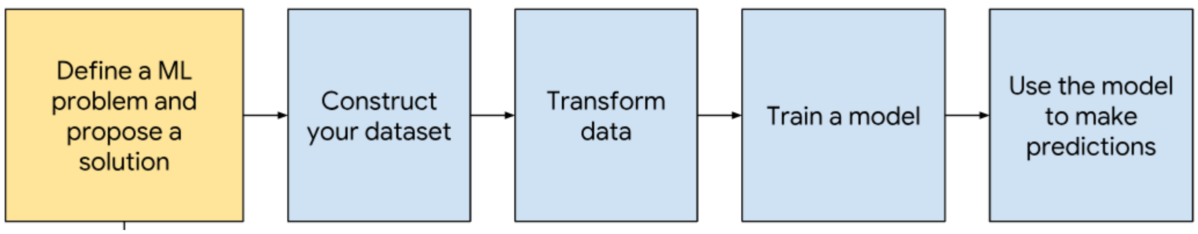


Figure 1. Concept of Machine Learning

(Source: https://www.sas.com/en_us/insights/analytics/machine-learning.html)

The process of machine learning starts by defining a problem that generally automation can improve but, in most cases, automated machine learning concepts are used to provide better and accurate results. At this point the problem defined needs to specify the direct goals without concerning the indirect goals. The solution for the problem is also defined in the first phase itself and is then set as a goal to be achieved. Second part of the process is to select and transform data according to the chosen model so that the data can be fed to the model which is the most important part of the process. To do so useful data is identified and collected, from that data important feature extraction takes place providing us with a sampling strategy according to which the data is split into smaller chunks of useful dataset. The data fed to model is then used to train the model to perform predictions. The trained model is then fed a test dataset to check the accuracy of the model and the results of the predictions are analyzed. Figure 2 shows a the concept which is followed to train and test a model.

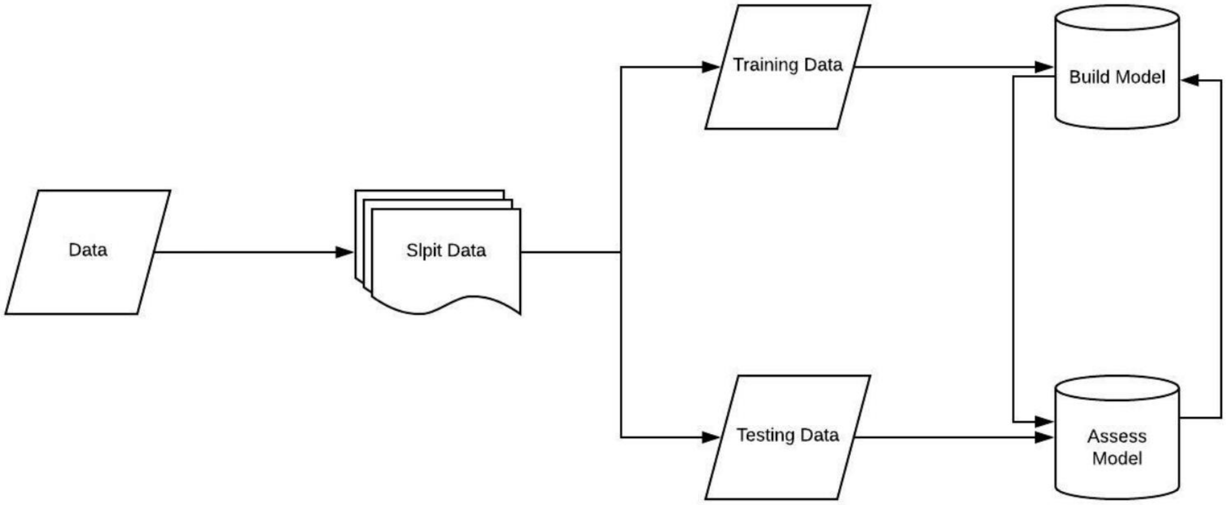


Figure 2. Training and Testing a Model

Python was used in this machine learning venture as Python is an ideally suited language for machine learning. This nature of Python stems from its readability and math-like syntax. Semantically, Python code resembles common mathematical ideas and its innumerable libraries offer many special functions and data structures that are useful for machine learning.

Types of Machine Learning:

Problems:

Regression

In regression, an exact precision is sought from the machine learning model. Here, training data is used to estimate a relationship between entities by fitting it into an equation (e.g. forming a best fit line). This relationship or equation is then used to make predictions.

Classification

In classification, the requirement is to assimilate data into categories and use this to make predictions about which category a certain input will belong to. It is not essential to have pre-classified data to solve this problem.

Solutions:

There are 3 typical ways of solving the problems depicted in the Figure 3, namely:

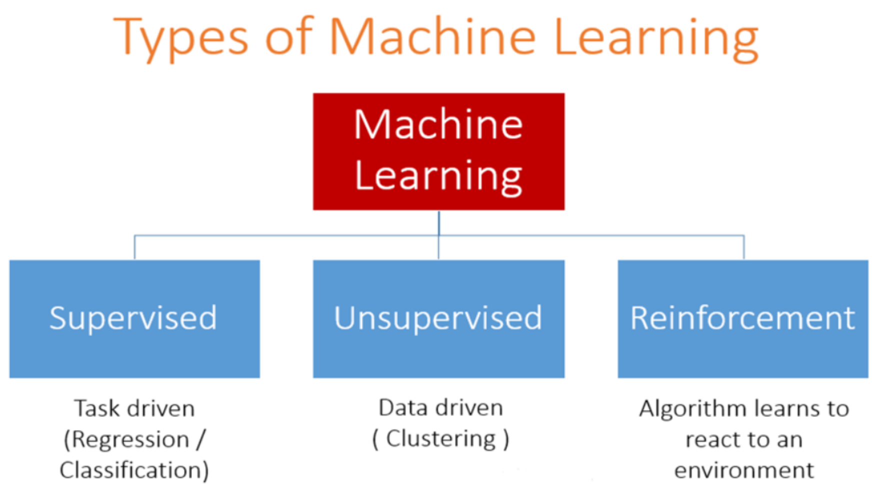


Figure 3. Types of Machine Learning

(Source: https://www.analyticsvidhya.com/blog/2015/06/machine-learning-basics/)

Supervised Learning:

Supervised learning is a technique of machine learning that is used when there is labelled training data to train the model on. In this form of learning, the data is converted into a vector that is given as input to learning functions which then map this data to a model. The model is then used to make predictions. This form of machine learning can be used for both regression and classification problems.

Unsupervised Learning:

This technique of machine learning is used in situations where there is no labelled training data. Here we do not know the correct mappings of data to train models, but data is analyzed for commonalities and clustered based on similarity and difference. This form of machine learning can only be used for classification problems.

Reinforcement Learning:

Reinforcement learning is a technique that is generally used in game theory and control theory. Here, the system tries to optimize itself towards a cumulative reward by constantly seeking reinforcement on the actions it takes. It can adapt easily to different scenarios where typical models would otherwise fail.

At SAP Labs, a large set of pre-classified data was supplied for a classification (labelling) problem. This data was raw and contained a large amount of noise.

The raw data at SAP was photos of all the products available in the SAP catalogue that is available to all its customers. This catalog is what a strategic purchaser uses to decide what materials go into a contract. The system that was required was one that could help recommend products from this catalog to the strategic purchaser.

3.1 Building Data visualizations

In order to better understand the data, the first task that was undertaken was building visualizations of the data that was provided. Because of the large amount of training data, a mechanism capable of handling mass amounts of data was necessary and the ELK stack presented as a likely candidate for this analysis.

The ELK stack consists of three components: Elasticsearch, Logstash and Kibana. Elasticsearch is a highly distributed, sharded database that can handle a colossal amount of data. Logstash is an Extract, Transform and Load (ETL) tool that takes a pattern it needs to search for in raw data, extracts useful information based on this pattern and stores it in Elasticsearch. Kibana is a Visualizer that is based on an Elasticsearch database. On setting up a Kibana visualizer for the Elasticsearch database, different visualizations were configured and obtained in an instant that showed the representation of different features in the raw data.

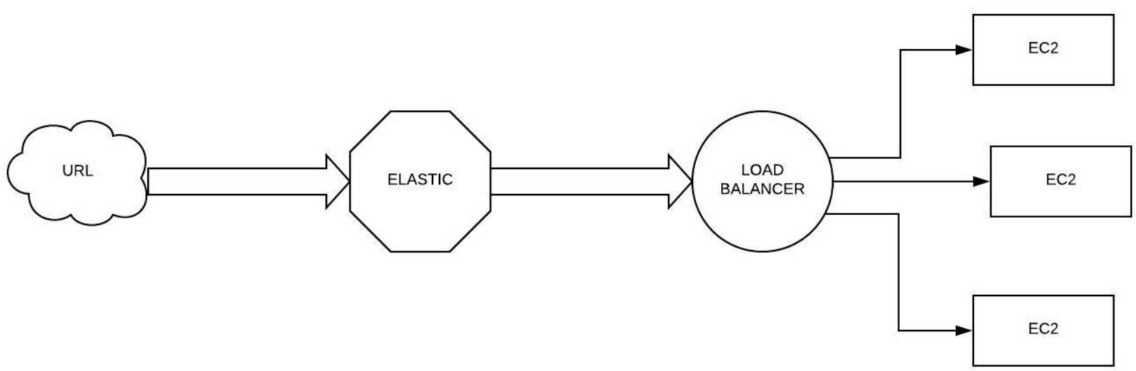


Figure 4. AWS EC2

When the first visualizations were built, it became apparent that there was a lot of noise in the data and it was proving to be difficult to draw any reasonable conclusions. The noise, however, was quickly identified to be coming from items in the background of images, unclear images, image flares, shadows etc. The visualizations thus helped identify noisy sections and prune them out of the dataset. Additionally, an attempt was made to obtain a cleaner dataset from labelled datasets obtained from scraping google images just for a comparison of the model's performance on it.

Upon cleaning the data, the visualizations became far more meaningful. From these visualizations we were able to draw reasonable conclusions about how to further clean this data, what features were important for extraction and what models were likely candidates for the data at hand.

3.2 Gathering required data and feature extraction

To analyze pictures, they were first converted into a relevant histogram containing pixel wise color information about the data. To do this, the python library skimage was used. Skimage is a python library that is used commonly in machine learning problems involving images. The histograms obtained from these images are obtained as Numpy sparse matrices. Numpy is a python library that is commonly used to solve numerical problems.

The numpy sparse matrix contains is a two-dimensional array where each column represents a feature in the image and each row represents a document (which in this case is an image). Edge detection, corner detection, blob detection and ridge detection along with thresholding were used as feature extraction mechanisms to obtain cleaner data for processing. Using these mechanisms, only certain parts of the images that store information important to all shapes within the image are retained and the other information is removed from the numpy matrix.

Shadows proved to be the most difficult part of the image to identify and prune given they formed shapes identical to that of the image. Finally, a mechanism was developed that identified shadows based on color and shape similarity to the main object in the image.

After performing all the transformations, the cleaned numpy array was stored and made readily available to all members of the team of data scientists working to develop models with it.

3.3 Selecting the right model

SAP uses numpy and multiple internal Python libraries in order to implement their patented machine learning algorithms. Using these libraries, 7 different models were constructed by the team of data scientists.

Three of these models were Convolutional Neural Networks (CNN) that was built with Keras. Internally, they used TensorFlow and ran on specialized hardware made by Nvidia and located at their Santa Clara office. SAP shares a business partnership with Nvidia given their proficiency in making graphic cards that are required in machine learning activity.

The remaining four models were built using ensemble methods that combined the CNNs in the order shown in Table 2.

Table 2 Ensemble Models

Presence matrix	CNN1 – Model 1	CNN2 – Model 2	CNN3 – Model 3
Ensemble 1 - Model 4	Y	Y	N
Ensemble 2 – Model 5	Y	N	Y
Ensemble 3 – Model 6	N	Y	Y
Ensemble 4 – Model 7	Y	Y	Y

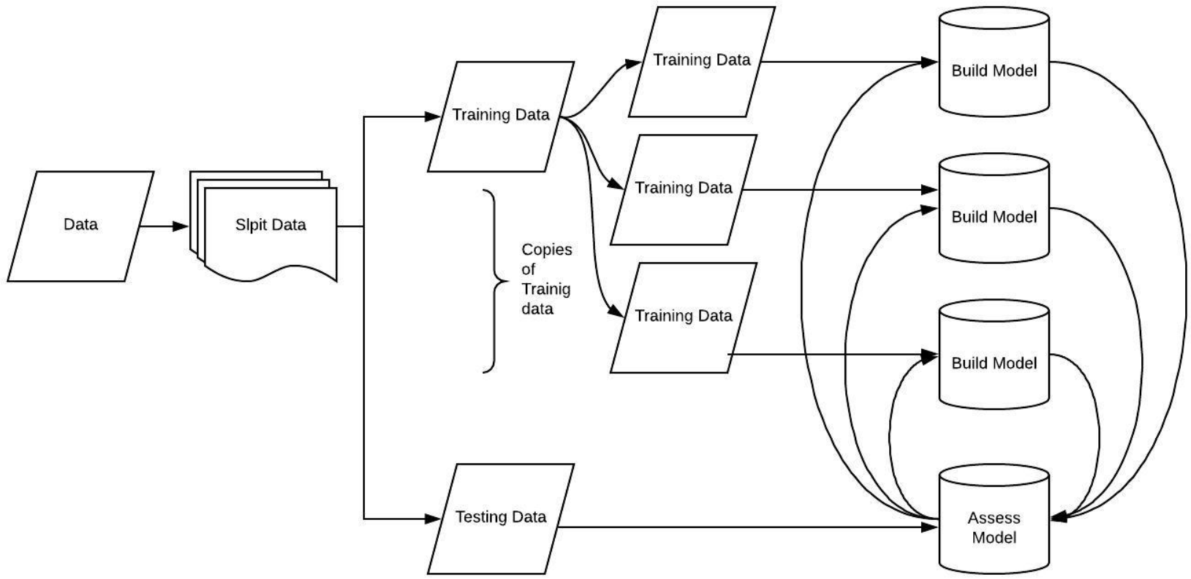


Figure 5. Training the Model at SAP

Post construction of the code, the hardware was procured to run up to 3 models at once. Each model was run with the same training data.

3.4 Training and Testing

SAP has a customer base of 330,000 businesses worldwide. The procurement data of over 200,000 classes (and even more sub-classes) of products for all these businesses for the last 10 years were available to train these models.

The cleaned pre-classified dataset was split into two parts: Training data and Testing data. Training data is the data which is used to train the model and constitutes 80% of the total pre-classified dataset. Models that are trained on this training data are tested against the remaining

20% testing data. The results obtained from the model are compared with the pre-classified information to draw a metric on how well the model performs.

In order to train multiple models simultaneously, multiple copies were made of the training data. Then 3 models were simultaneously supplied with the training data shown in Figure 5 and run on different hardware under different threads. The three models that were trained were the three CNNs. Following training each model, they were tested by running them against the testing data as though it was unclassified. The classified results obtained from the model were then compared against the actual results in the testing data.

The results obtained from the CNNs were then combined to build the results of the ensemble methods as shown in the Figure 6.

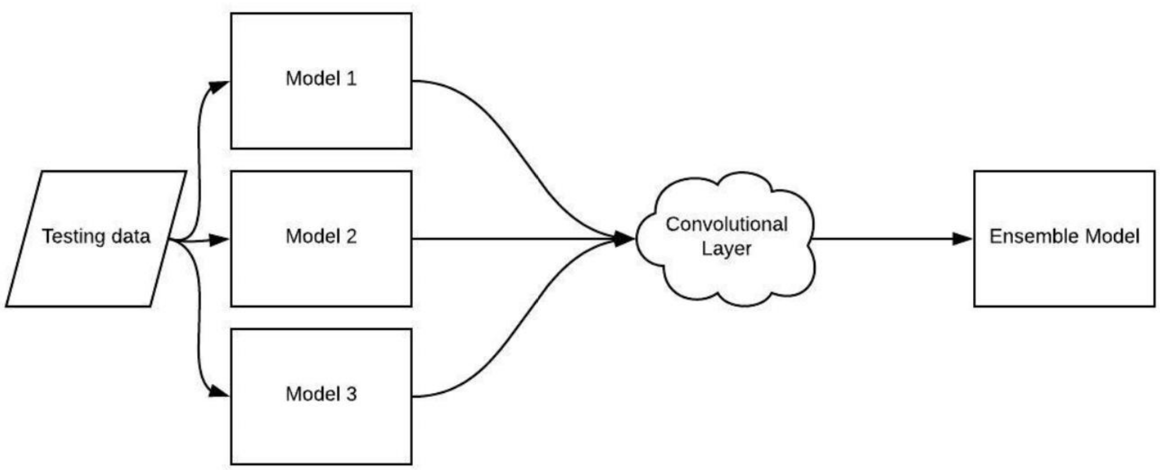


Figure 6. Ensemble Model built using CNN

3.5 Model Tuning

Each model takes multiple different parameters like the depth of the CNN, number of layers, number of iterations etc. Multiple variations of these parameters were tried and the CNNs were trained to give the best responses. Each time a model was re-run, the ensemble methods tied to that model were also run to update the ensemble results. Each method was compared separately although the maximization of the various metrics (to be discussed later) were focused on the CNNs with the assumption that better performing CNNs will combine to form better ensemble methods.

These local maxima's for each CNN were hard to achieve and one model took 30 iterations before arriving at its optimal best. Multiple tweaked versions of the same model were trained simultaneously until the best performer was obtained.

Once the best performing CNNs were obtained, the weights of the ensemble methods were varied to produce different results. While this was easier to do as the CNNs did not need to be retrained, it still took a significant amount of processing effort and time to get predictions for the testing data for each variation of each ensemble method. A feedback mechanism was also built which used the metrics obtained from each trained version of the model and generated new sets of weights and balances for the next iteration which it predicted would give better results. The feedback mechanism further ensured that the best possible results will be obtained from each ensemble method. The data scientists could also intervene in the feedback loop to reject models. Every rejection compulsorily required a statistical reason which the model the used to get better predictions for the next iteration of weights.

The result was a total of seven models working at their tuned best and giving the best accuracy achievable with the supplied training data.

3.6 Analysis

The business standards that were set for the accuracy of the model was 88% given SAPs high standards and prestigious customer base. Individually, the Neural Networks were found to reach up to 82% accuracy.

One model reached up to 96% accuracy on the testing data but was found to be a little under 68% for live data which shows that the model was overfitted from the tuning that was performed on it. These overfitted models were required to be re-run with different parameters. The ensemble methods that incorporated these models were also affected in this process.

The ensemble methods were however, far superior to the individual CNNs and reached up to 94% on testing data and 90% on live data – clearing the thresholds to be a viable product option. This cemented the idea that no one model could be used in all scenarios, but a combination of different models could be optimized to work well for a specific scenario.

Additionally, data scraped from Google images was also run against the set of models (both to train and test). It was found that because of better clarity in the images, the models did better with the average accuracy of models showing a significant increase. However, due to proprietary concerns, the models could not be trained on this dataset.

While some ensemble models did do a lot better than individual CNNs, a lot of them did worse and tuning ensembles turned out to be far more cumbersome than training CNNs.

4. Performance analysis of built systems

Systems that are built at conglomerates like SAP that service such many customers, are required to perform at a high standard. These standards include both speed and accuracy of responses and can be quantified by many techniques. In order to evaluate these standards, it is imperative to run the built system against various benchmarking tests and drawing performance statistics from them.

Apart from this, scalability, availability and consistency are of importance in products at SAP. The product that is built must be able to service a large customer base while always being available with minimal downtime. At the same time, it must maintain a degree of consistency with its predictions. The system was to be scalable for future use. Performance being business critical in most of the cases is a deal breaker if certain standards are not met.

4.1 Performance metrics

Metrics used to measure an organization's activities, performance and behavior are generally known as the Performance Metrics. It gives a bigger picture about how the organization is performing and how well the resources in the organization are managed. How well the workers of the organization are performing their tasks.

A key performance index (KPI) is the measure of how well the organization is able to meet its key objectives and how well are they able to meet their targets. Performance Metrics done were-

1. Confusion Matrix: It is one of the easiest and intuitive metrics used to find the accuracy and the efficiency of the model. It is generally used for classification problem where the output might me of two or more different classes. In the matrix the actual classifications are represented as columns and the predicted ones are the represented by the rows. Figure 7 shows a Confusion Matrix.

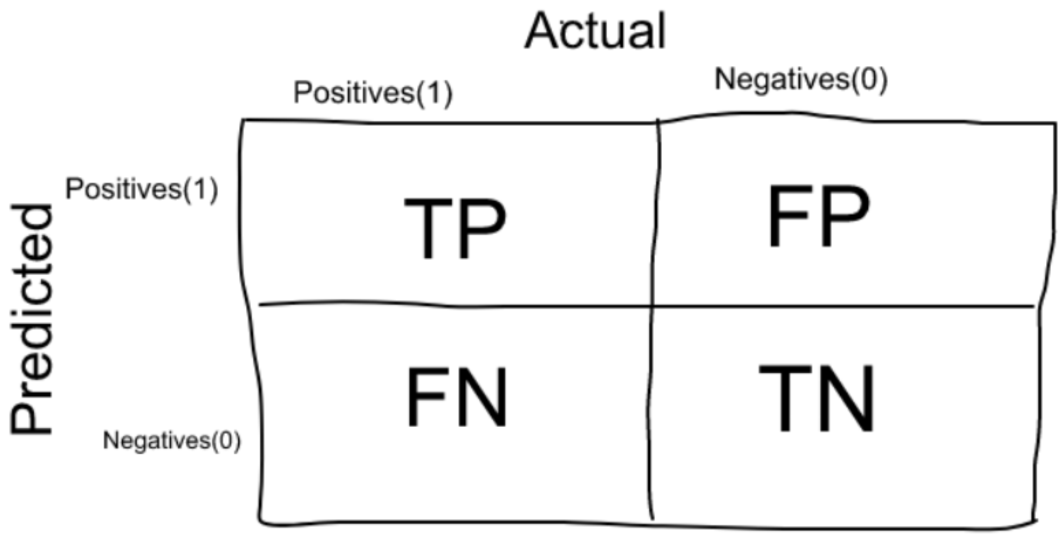


Figure 7. Confusion Matrix

Terms associated with a confusion matrix-

- True Positives (TP)- Cases where the actual points were true, and the predicted points were also true.
- True Negatives (TN)- Cases where the actual points were false, and the predicted point is also false.
- False positive (FP)- Cases when the actual points were false, but the predicted points are true. It's called false positive because the model prediction was wrong hence false but was predicte true by the model thus positive.
- False Negative (FN)- Cases when the actual point was true but was predicted by the model as false. It's called false negative because the model prediction was false, and it predicted a false hence the negative.

For an ideal scenario we would want our model to give zero False Positives (FP) and zero False Negatives (FN), which is not true in any case in real world. Thus we try to achieve maximum efficiency by reducing these.

2. Accuracy: the ratio of total number of correct predictions made to the total number of predictions made.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

In the Numerator, are our correct predictions (True positives and True Negatives)(Marked as red in the Figure 8 below) and in the denominator, are the kind of all predictions made by the algorithm(Right as well as wrong ones).

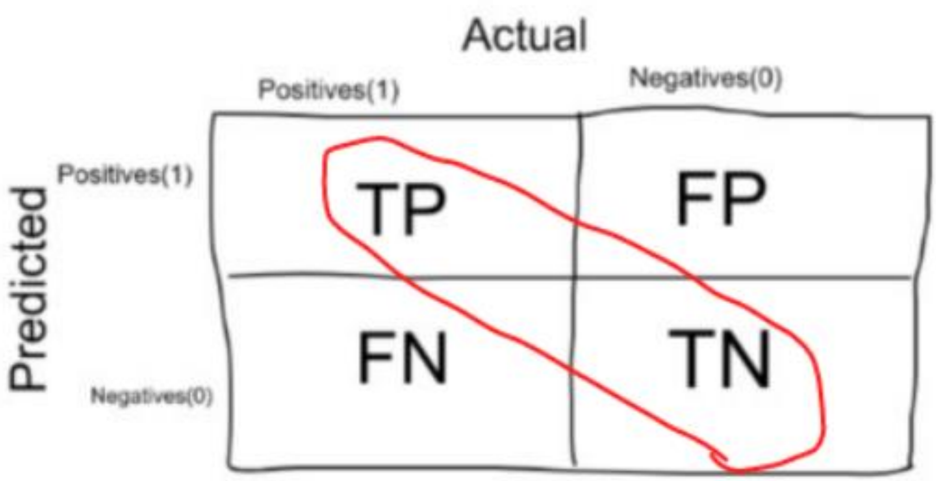


Figure 8. Accuracy

For balanced classes accuracy is a good measure, it should never be used if the targeted classes have majority of one class.

3. Precision: The frequency at which the model was able to predict any points correctly is the precision of the model. Shown in Figure 9.

$$\text{Precision} = \frac{TP}{TP+FP}$$

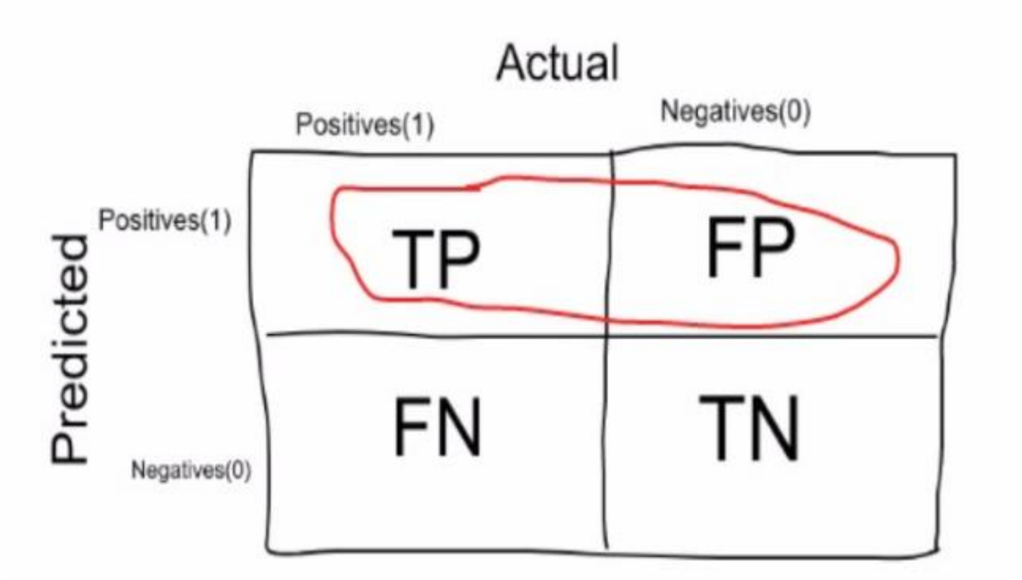


Figure 9. Precision

4. Recall or Sensitivity: The ratio between the number of possible positive points to the number of points that the model predicted correctly as positive is the recall. See Figure 10.

$$\text{Recall} = \frac{TP}{TP+FN}$$

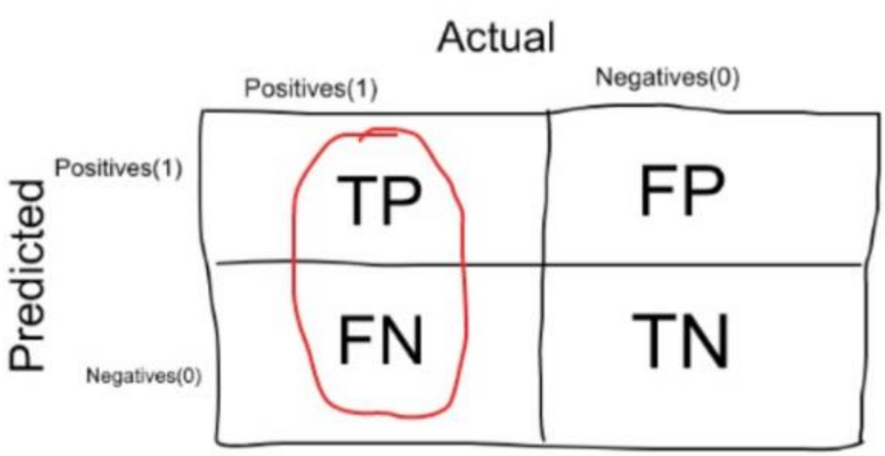


Figure 10. Recall or Sensitivity

So, basically, if we want to concentrate more on minimizing false negatives, we want our recall to be as close as possible to 100% without precision being too bad and if we want to concentrate on minimizing false positives, then our focus should be to make precision as close as possible to 100%.

5. Specificity: Its is the opposite of recall. The ratio between the number of possible negative points to the number of points that the model predicted correctly as negative.

   $\text{Specificity} = \frac{TN}{TN+FP}$ See Figure 11.

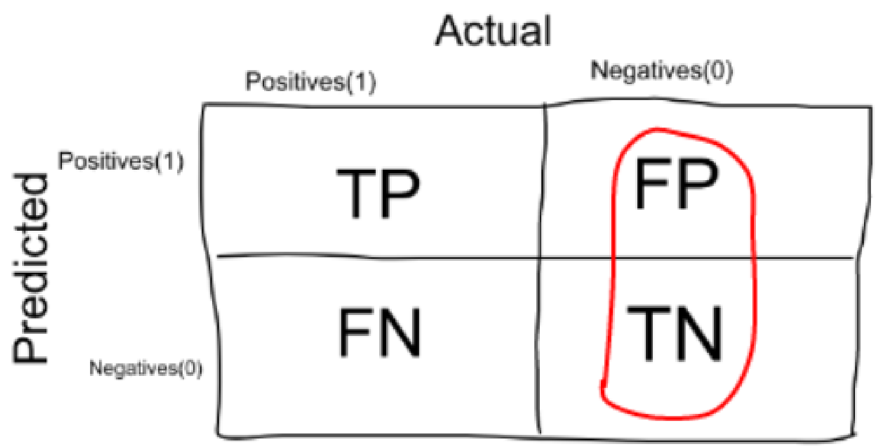


Figure 11. Specificity

6. F1 score: We want a score that can represent both Precision and Recall. Simple way to do this is by just taking the arithmetic mean of them. Which turns out to be bad in most cases. Thus, we need something more balanced than the arithmetic mean such as a harmonic mean.

Harmonic mean is given by the formula shown in Figure 12 below:

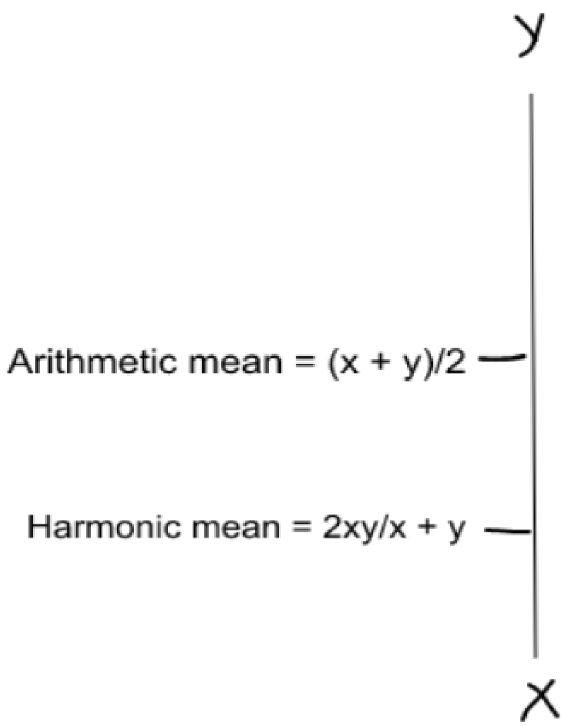


Figure 12. Harmonic Mean

Harmonic mean is the arithmetic mean when both x and y are equal. But when they are different than the mean shifts closer to the smaller number as compared than larger number.So if one number is really small between precision and recall, the F1 Score kind of raises a flag and is more closer to the smaller number giving the model an appropriate score rather than just an arithmetic mean.

4.2 Gathering performance data

To run the performance analysis, the system was kept as close to its production reality as possible. In order to achieve this, they were run on high-performance computing machines (HPC machines) at Nvidia. These high-performance computing machines were provisioned using virtual machines that were accessible like an IaaS. A web application was used to provision the virtual machines. After they were spawned and details concerning their IP addresses were gathered. Connections were established to these HPCs from desktops and the process intensive code was run there.

Some such VMs were put into a suspended state and copied to create live machine instances that could be spawned in an instant. These suspended machine images were provided to a load balancing mechanism which then exposed a port. This port was secured by the firewall and was only accessible from internal SAP networks unless explicitly granted permission.

The number of VMs running the code was set to vary between 20 and 50 based on load. When a request is hit, the load balancer will use a balancing algorithm to decide which VM is best equipped to service this request. If all VMs hit a certain threshold of activity, the number of VMs was increased. Similarly, when the VM activity fell below a certain threshold, it was terminated. The number of VMs never dropped below 20 and never increased above 50.

4.3 Making an API

The end point obtained from this system was then tested and analyzed for its performance. To test it, a variety of tests were run to simulate different conditions like high load (many requests), high stress (large volume of data in requests), bad network conditions etc. Data for the performance analysis came from VM statistics and BlazeMeter.VM statistics were obtained from the IaaS provider and BlazeMeter is a performance testing DevOps tool made by Apache and can gather performance statistics and building live visualizations of the same. These tests also gave more information on the scaling numbers that were needed for the application.

In addition to this, the load balancer also periodically checked all running VMs for a heartbeat to ensure that they were running as expected. If an anomaly in the VM stats exceeded a certain threshold or persisted beyond a certain amount of time, the VM was killed and the request was transferred to another VM while spawning a replacement. The same request would, however, not be serviced by more than 3 VMs i.e. the load balancer would give up on servicing a request after 3 tries and send an HTTP 5XX error to the client.

4.4 Visualizing information gathered

In the final sprint of the product pre-release, it was packaged and provided to a central server team at SAP for further extensive testing and code reviews. Once it was approved by this team, the product was ready to be shipped.

The central team pushes the packaged feature to the SAP business solutions bundle server and all customers are then prompted to update their product to then use this new feature. Data obtained after customers update the product is live data which the system was set up to capture and send as a pipeline to various charts made using pandas. This live data provided deep insights into how the model is performing with live data. While most models performed consistent to their performance on training data, some models did significantly worse on account of being overfitted – this was easily seen in the visualizations.

Real life statistics of the end point were also monitored. This includes wait time, blocked time, DNS wait time, connect time, send time, receive time and SSL wait time. In addition to this Round-Trip Time was also monitored for requests going to this system. This provided information to ensure the system was running and servicing requests in time. This data was further used to tune the load balancer.

Live usage data was monitored to see the overall effectiveness of the product. While the acceptance rate of the recommendations provided by the system were monitored, customers were provided the option to provide feedback about recommendations that were rejected. This data was again used to tune the model.

5. Recommendation

- Maintaining excellent documentation is a must. All the teams need to document the different plans and changes made to the existing system.
- Data collected can be fed for unsupervised learning and can be start of a neural network built on it.
- Recent introduction of integrated machine learning in the AWS EC2 which trains from the stats that are recorded can be used to create better dashboards and can help us understand more features about the data.
- Better communication between the cross functional teams to reduce duplication of work.
- Having virtual meeting with the customers to make them completely understand the improved functionality of the product.
- Better labeling of data will help better feature extraction.
- Recommending customer to make the contracts in advance rather than keeping it till the last moment facilitating better decisions in future.

6. Conclusion

This project illustrated the necessity of a product to effectively and beneficially helps the business during the process of procurement of materials. Gap analysis listed out the main points that needed to be focused on and how they were to be tackled. Requirements gathering, and intensive analysis helped to figure out how to proceed with the task at hand and the team decided to implement Machine Learning for making the process better. Multiples models trained and tested giving the team the best ensemble model suited for the job. Performance metrics were used to compare the models and deciding on which model performed the best. The improved procurement assistant was beta released and tested internally and was a success exceeding expectations of the management.

A constant feedback and self-improvement mechanism ensure that the product will last in future and is constantly improving based on the recommendations given by the users. It will help improve relations between the supplier and the purchaser. Maximizing the profitability, reducing the waste and providing the ability to choose the maximum beneficial opportunity for an organization was the main aim of the product which was achieved beautifully using the power of analytics, machine learning and hard work.

References:

Sparse Matrix and its representations | Set 1 (Using Arrays and Linked Lists). (2017, September 10). Retrieved August 08, 2018, from https://www.geeksforgeeks.org/sparse-matrix-representation/

Shung, K. P. (2018, March 15). Accuracy, Precision, Recall or F1? – Towards Data Science. Retrieved August 10, 2018, from https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

What Is Amazon EC2? (n.d.). Retrieved August 15, 2018, from https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html

Machine Learning: What it is and why it matters. (n.d.). Retrieved August 16, 2018, from https://www.sas.com/en_us/insights/analytics/machine-learning.html

Machine Learning Glossary| Google Developers. (n.d.). Retrieved September 01, 2018, from https://developers.google.com/machine-learning/glossary/

Education – Google AI. (n.d.). Retrieved September 01, 2018, from https://ai.google/education/

Rules of Machine Learning: | ML Universal Guides | Google Developers. (n.d.). Retrieved September 05, 2018, from https://developers.google.com/machine-learning/guides/rules-of-ml/

Reinforcement learning. (2018, December 04). Retrieved December 04, 2018, from https://en.wikipedia.org/wiki/Reinforcement_learning

SAP Labs | About SAP SE. (n.d.). Retrieved October 18, 2018, from https://www.sap.com/corporate/en/company/innovation/sap-labs.html

Statistical classification. (2018, October 28). Retrieved November 2, 2018, from https://en.wikipedia.org/wiki/Statistical_classification

What is gap analysis? - Definition from WhatIs.com. (n.d.). Retrieved November 05, 2018, from https://searchcio.techtarget.com/definition/gap-analysis

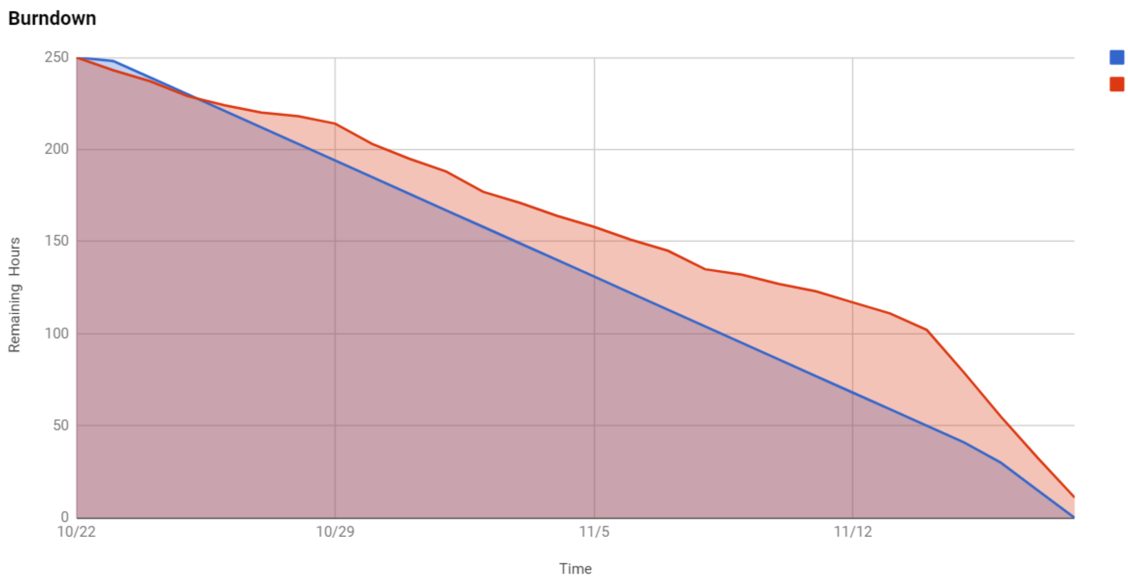
Why is Python so popular in machine learning? (n.d.). Retrieved November 05, 2018, from https://www.techopedia.com/why-is-python-so-popular-in-machine-learning/7/32881

Ouaknine, A. (2018, January 16). Review of Deep Learning Algorithms for Image Classification. Retrieved November 11, 2018, from https://medium.com/comet-app/review-of-deep-learning-algorithms-for-image-classification-5fdbca4a05e2

ImageNet: VGGNet, ResNet, Inception, and Xception with Keras. (2017, September 28). Retrieved November 12, 2018, from https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/

Sunasra, M. (2017, November 11). Performance Metrics for Classification problems in Machine Learning. Retrieved November 14, 2018, from https://medium.com/greyatom/performance-metrics-for-classification-problems-in-machine-learning-part-i-b085d432082b

Appendix:



Burndown chart for the team