

**HINDUSTAN AERONAUTICS LIMITED**

**AIRCRAFT DIVISION**

**NASHIK**

TRAINING PROJECT REPORT

**“AUTOMATIC SNAG DISPOSITION AND MANAGEMENT SYSTEM USING M.L”**

Mr. MOOLA VISHWACHANDRA RAJENDER

Mr. MEDIBOYINA RUPESH KUMAR YADAV

B.Tech (CSE) 2ND YEAR

NATIONAL INSTITUTE OF TECHNOLOGY

ANDHRA PRADESH

PROJECT GUIDE:

Mr. M. RAJENDER

DGM (DESIGN)

**AIRCRAFT UPGRADE RESEARCH AND DESIGN CENTRE**

**CERTIFICATE**

This is to certify that Mr. MOOLA VISHWACHANDRA RAJENDER, Roll No. 420210, Mr. MEDIBOYINA RUPESH KUMAR YADAV, Roll No. 420207, of B.TECH. CSE 2nd Year from NATIONAL INSTITUTE OF TECHNOLOGY, ANDHRA PRADESH, has successfully completed training and project work during the period 2nd June to 3rd July 2022 on ***“AUTOMATIC SNAG DISPOSITION AND MANAGEMENT SYSTEM USING M.L”*** at Design Quality department at Aircraft Upgrade Research and Design Centre (AURDC), Hindustan Aeronautics Limited (HAL), Nasik, as a part of Industrial Internship.

PROJECT GUIDE HEAD OF DEPARTMENT

Mr. M. RAJENDER Mr. M. RAJENDER

DY. GENERAL MANAGER DY. GENERAL MANAGER

(DESIGN) (DESIGN)

# **ACKNOWLEDGEMENT**

This is the report of the project I undertook during Internship for duration of 4 weeks at Hindustan Aeronautics Limited, Nasik.

I would like to **thank MR. S. J. BHOLE, GENERAL MANAGER (AURDC)** HAL, Nasik for giving me this excellent opportunity.

I would like to express my gratitude to my project guide **Mr. M. RAJENDER, DY. GENERAL MANAGER (DESIGN)** who has guided me through each phase of my project. His guidance, suggestions and invaluably problem solving and friendly advice during our discussions, are much focused to give me a proper direction to analyze and understand the project objective.

I am grateful to **TRAINING DEPARTMENT** for giving me the opportunity to undertake the project work at “M/s Hindustan Aeronautics Limited, Aircraft Division, Nasik, India”.

I would also like to extend my sincere thanks to all the staff members and employees.

**ABSTRACT**

The Aircraft Manufacturing and Overhaul is a complex and technical intensive job. At various stages of aircraft manufacturing and overhaul faults do occur due to various reasons. The faults referred as SNAGS are reported by Quality to Design department for the disposition of the same. The Design Engineers study the problem and dispose accordingly. The disposition is in terms of acceptable, not acceptable and rework depending on the criticality of the components/system.

Since this is monotonous and time consuming sometimes takes very long time for the disposition of snags. It is proposed to implement Machine Learning model to automatic disposal of snags. The data has been gathered over 13 years is available. Since data is available for the snags and their dispositions, it is feasible to adopt Machine Learning/Artificial Intelligence for the same.

This project is an implementation of Various Machine Learning Algorithms. This project aims to develop, train, test and predict whether the detected SNAG (Fault) is Acceptable to Design, Not Acceptable to Design or Rework is required based on the provided inputs from the user.

**Table of Contents**

[**ACKNOWLEDGEMENT** II](#_Toc68350497)

[**ABSTRACT** II](#_Toc68350498)

[**ABOUT COMPANY** 2](#_Toc68350499)

[**INTRODUCTION** 2](#_Toc68350500)

[**SYSTEM CONFIGURATION, SOFTWARE AND EQUIPMENT** 2](#_Toc68350501)

[**PROJECT** 2](#_Toc68350502)

[1. Problem Statement 2](#_Toc68350503)

[2. Data Analysis 2](#_Toc68350504)

[3. Data Preprocessing 2](#_Toc68350505)2

[4. DATA Preparation 2](#_Toc68350506)6

[5. Machine Learning Algorithm 2](#_Toc68350507)6

[6. Model Development 2](#_Toc68350508)7

[7. Evaluation](#_Toc68350509) 20

[8. Model Fine Tuning 2](#_Toc68350510)2

[9. Final Model Results 2](#_Toc68350511)4

[10. Results 2](#_Toc68350513)7

[**CONCLUSION** 2](#_Toc68350514)7

[**REFERENCES** 2](#_Toc68350515)9

# **ABOUT COMPANY**

Hindustan Aeronautics Limited (HAL) is a premier Aerospace company in Asia which is engaged in Design, Development and Manufacture of military and civil aircraft for over 75 years. It was established as Hindustan Aircraft in Bangalore in 1940 by Seth Walchand Hirachand to produce military aircraft for Indian Air force.

HAL is an organization where integrated airborne systems in the form of fighter aircraft and helicopters are conceived, developed, manufactured and serviced. It is one of the few corporate giants in Asia whose capabilities span the entire range of activities from product conception to after sale support. HAL is also involved in manufacture and assembly of structure required for India’s Space programs.

Hindustan Aeronautics Limited is the largest Public Sector Unit (PSU) under Ministry of Defense and is a Navratna Company. The company takes up maintenance and overhaul services to cover the life cycle requirement of all old and new products. Presently, 13 types of aircraft/helicopters and 17 types of engines are being overhauled. Additionally, facilities for repair/overhaul of various accessories and avionics integrated on aircraft of Russian, Western and Indigenous designs are also provided. HAL has 20 production Divisions, 10 Research and Development Centers and one Facility Management Division.

HAL Nasik has 3 divisions:

* Aircraft Manufacturing Division (AMD)
* Aircraft Overhaul Division (AOD)
* Aircraft Upgrade Research and Design Centre (AURDC)

Aircraft Division Nasik, established in the year 1964 for license manufacture of MiG-21FL aircraft & K-13 Missiles, is located at Ojhar, 24 kilometers from Nasik and approximately 200 kilometers from Mumbai in the state of Maharashtra. The division since then manufactured other MiG variants; viz MiG-21M, MiG-21 BIS, MiG-27 M and the state-of-the-art aircraft i.e. Su-30 MKI. Along with manufacturing, the division also carries out overhaul of the MiG series aircraft and started ROH of Su-30 MKI.

With the introduction of New a state-of-the art project i.e. Su-30 MKI and for smooth activities and operation of the division then  Aircraft  Division, Nasik is de-lineated into two Divisions viz Aircraft Manufacturing Division (AMD) for manufacturing activities  and Aircraft Overhaul Division (AOD) for Repair & Overhauling (ROH) activities.

|  |  |
| --- | --- |
|  |  |
| **SU-30MKI** | **SU-30MKI Touchdown** |

The Division is currently engaged in the following activities:

* Manufacturing of Su-30 MKI Aircraft.
* Supply of spares / units /consumables and aerospace fasteners.
* Support to HAL and Non-HAL made aircraft of Russian Origin in terms of Life extension, Modifications, Site Repairs and investigations.
* Diversification into Civil Aircraft manufacture by taking up work packages.
* Export of MiG spares to Egypt, Syria, Vietnam, Malaysia, Algeria, Poland & Russia.

|  |  |
| --- | --- |
| Russian official confirms talks to sell 21 MiG-29 fighters to IAF - The Week |  |
| **MiG 29** | **MiG 27 flying in Formation** |

**AURDC Division:**

* AURDC was established as Design Department in 1964 to provide design support to the manufacturing program, it has grown into a full-fledged R&D center named Aircraft Upgrade Research and Design Centre (AURDC). With vast experience on Su30MKI, MiG series aircraft in design and development for product improvements, role capability enhancement, indigenization, technology up-gradation, structural integrity studies for life extension, flight test analysis and mid-life upgrade are its strengths.
* AURDC over the years gained experience in design and development of new airborne system, Integration of systems and weapons in the aircraft. Evolved many modifications in terms of reliability, maintainability, operation, capability, enhancement and etc.

# **INTRODUCTION**

Aircraft manufacturing and overhauling is a diverse and vast process. In each step of the process, we need to make sure the quality and standards are maintained with utmost importance. Furthermore, each part or equipment of the aircraft is manufactured locally at the Aircraft Manufacturing Division (AMD) of HAL, Nasik, or has to be sourced from another HAL manufacturing facilities, and some specialized licensed parts are also sourced from. With this level of the abstruse, multistage process it fairly regular for minute things to go wrong.

These defects or snags in more technical term, needs to be taken care of. These snags are generally small faults but to high intricacy and importance about the quality of these aircrafts, they can’t be ignored. Each snag often as minute as an error of 0.1 mm as reported to the quality control.

Due to manufacturing and testing being a hand-in-hand process for aircraft manufacturing, Snags reporting comes from various stages of the process from raw material/part-manufacturing/assembly shops to inspection.

Each of the reported snag is then sent to the appropriate department of Design for a full detailed analysis of the problem and to give appropriate Disposition. The Design department after completing the study of the snag disposes of the snag in one of three category 1) Acceptable to Design, 2) Not Acceptable to Design, or 3) Rework is required with a rework scheme.

Brief description of the same is given below **:**

1. Acceptable to Design - this means that the snag which was identified is not going to impact the safety & reliability/quality of the product.
2. Not Acceptable to Design – this means that the snag which was identified is going to be impending on the safety, reliability/quality and other aspects of the product.
3. Rework is required – this means that the snag which was identified is a minute problem and it can be fixed with some repair work. In such cases the Engineer from the designated Design department gives a solution to the snag and according to this disposition, rework is carried out.

The project aims to predict this classification based on different parameters that are related to the snag. These parameters mainly include Snag Description, Snag Stroke, Inspection Stage, Drawing no., and Part no.. This will result in a faster and efficient Disposition of snags without any need for major documentation and forwarding. And the rework snags will be forwarded to the engineer so that he can directly provide a solution instead of dealing with its classification first. This will hugely lower his workload and save it for higher priority R&D tasks.

# **SYSTEM CONFIGURATION, SOFTWARE AND EQUIPMENT**

PC Model: HP ZBook 17

Processor: Intel(R) Core(TM) i7-4600M 2 Core(s) 4 Logical Processor(s) @2.9GHz

RAM: 8 GB

OS: Windows 7 Professional 64-bit (6.1, Build 7601)

Display: 1920\*1080 (32 bit) (60Hz)

Software used:

* Anaconda Navigator/Prompt
* Jupyter Notebook

Programming Language:

* Python 3.8.5

Datasets:

A part of SQMS dataset that contains organized description all the snags that have occurred previously in the product manufacturing and overhaul.

# **PROJECT**

## Problem Statement

***To Design and Develop a Machine Learning Model, for the prediction of Snag Disposition based on SQMS (SNAG & QUERY MANAGEMENT SYSTEM) dataset. The live Snags taken for the validation of model.***

## Data Analysis

SQMS dataset is used as a base for this ML model. The SQMS dataset is a real-world dataset with the size of 10,000 unique entries of different snags that occurred at various stages of Production, Testing and at other Validation Stages .Each entry in this dataset was entered manually during the process of manufacturing and overhaul of a product/part.

Due to the fact that the dataset was created manually and with several people behind the creation of it, this dataset is full of typographical errors, wrong entries, proxies, spelling mistakes, null values, and many more errors. These inconsistencies in data needs to be fixed before the subjecting of data to the machine learning model.

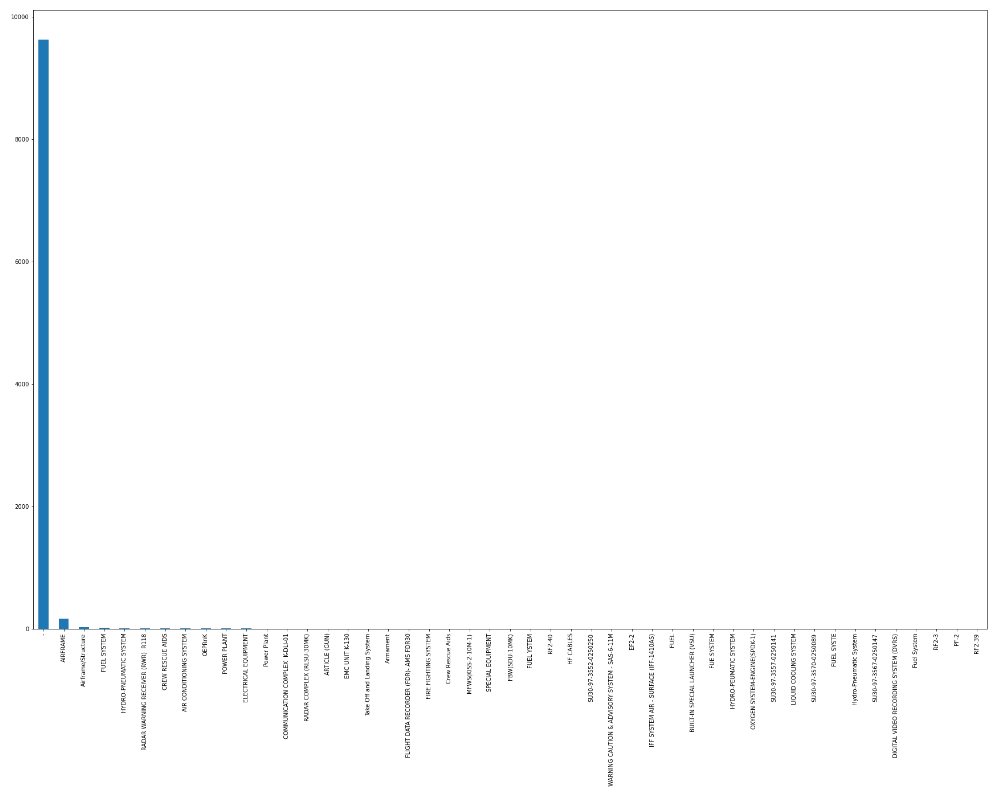
|  |  |
| --- | --- |
|  |  |

## Data Preprocessing

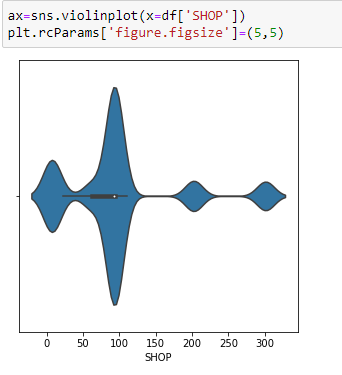
As the Data was collected from various stages, it is usual that many of them do not belong to specific designated system or sub-systems. Thus having Null entries.

**Parameters Null value counts**

* SNAG\_ID 0
* ACNO 3348
* INSP\_NAME 28
* SHOP 0
* INSP\_STAGE 273
* SNAG\_DATE 0
* SNAG\_DESC 0
* SNAG\_STROKE 13
* ENGR\_FLAG 962
* PART\_NO 6886
* TASK\_NO 4494
* SYSTEM 9632
* SUB\_SYSTEM 9635
* PROJECT 1608
* CLOSE\_DATE 3447
* DISPOSITION 112
* Forward Date 997
* Disp Date 123
* DWG\_NO 0
* With so many null values, the above parameters can’t be directly used for our model development and hence we need to process the data, if present extract the data from other columns.
* The rest of the parameters containing empty fields also needs to get located properly with some values, before model training.
* During data analysis, it is found that the dataset doesn’t contain any numerical value in terms of integer or float data type. So we replace all the empty fields with null values.
* The snag description and the disposition are the independent and dependent variables required for the existing problem statement.
* Snag description parameter contains additional data in it such as drawing number & part number. New parameter is derived in dataset named drawing number.



**Fig 2.1** The above figure shows the missing data present in System

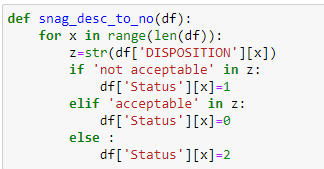


**Fig 2.2**

* A regular expression code is used to extract drawing number or part number, whichever is applicable for the particular snag and value is assigned to appropriate field in the dataset.
* The disposition is the dependent target variable in our proposed model. But value of this parameter varies for same class. This is mainly due to various different people involved in evolution of the dataset. Hence, we create a new parameter named ‘ Status ’ start assigning it values based on disposition.

0 - Acceptable to Design

1 - Not Acceptable to Design

 2 - Rework is required

**Fig 3.1**

Figure 3.1 shows code snippet used to convert the disposition into Status parameter.

* Finally all the spelling mistakes needed to be fixed before model training.



**Fig 3.2**

All the spelling mistakes had to be corrected manually. This correction process was by far the biggest task in the development of model.

## DATA Preparation

Once we had a good filtered dataset, we need to start preparing the dataset for model training and testing.

Since the independent variable values are text. It is imposed to use the word embedding’s and count vectoriser for the preparation of the data.

We append all the parameters into a single snag description parameter using pandas dataframe.

We also split the dataset in 85:15 ratio for training and testing respectively. For this purpose from **sklearn.model\_selection** library we used **train\_test\_split** function which is a standard in Machine Learning applications. We further declare various dataframe objects to store different parts of dataframe created through the function.

## Machine Learning Algorithm

Machine learning algorithms cannot work with raw text directly, the text must be converted into numbers. Specifically, vectors of numbers. Problem with modeling text is that it is messy, and techniques like machine learning algorithms prefer well defined fixed-length inputs and outputs.

In Natural language processing, the vectors x are derived from textual data, in order to reflect various linguistic properties of the text. This is called feature extraction or feature encoding .

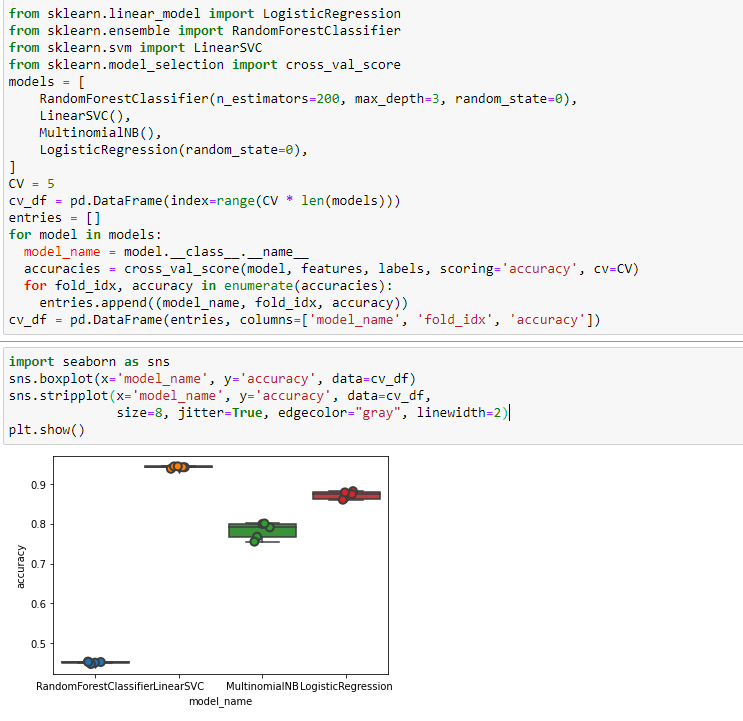
The Multinomial Naive Bayes algorithm is a Bayesian learning approach popular in Natural Language Processing (NLP). Multinomial Naive Bayes algorithm is suitable for Classification with Discrete Features and Support Vector Machine (SVM) , SVMs do not require any parameter tuning, since they can find good parameter settings automatically. All this makes SVMs a very promising and easy-to-use method for learning text classifiers

The biggest difference between the models you're building from a "features" point of view is that Naive Bayes treats them as independent, whereas SVM looks at the interactions between them to a certain degree, when using a non-linear kernel .

## Model Development

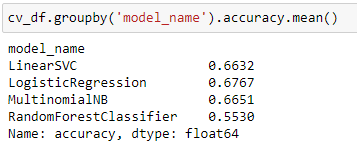
Various models were explored during the initial stages of the project. This was achieved by training and testing the models on default parameters of the model and then gradually changing parameters in an incremental manner to monitor the change in testing accuracy score, training accuracy score, Precision, recall, and fscore .

Undermentioned are the various testing results that were observed in the process:-



**Fig 4.1**

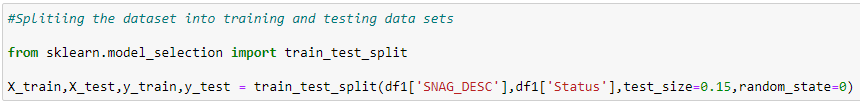
Fig 4.1 shows the results of 1st testing of data on different machine learning models.



**Fig 4.2**

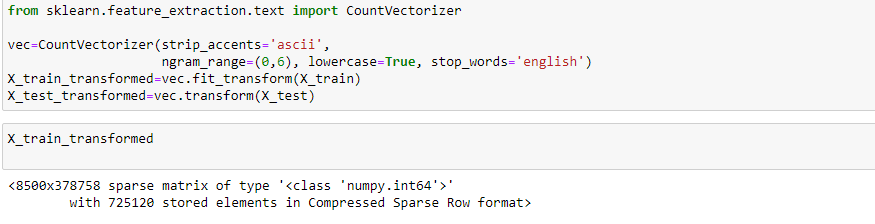
We can see that the Logistic Regression model has the accuracy score of 67.7% but Logistic Regression being a binary classifier it can’t be used to classify multiclass targets. And this explains why Not Acceptable has 0 value for all different types of scores.

Based on the accuracy and all other scores we come to a conclusion that Support vector Machine using non-linear kernel and Naïve Bayes’ MultinomialNB model is most suited for our classification requirements.

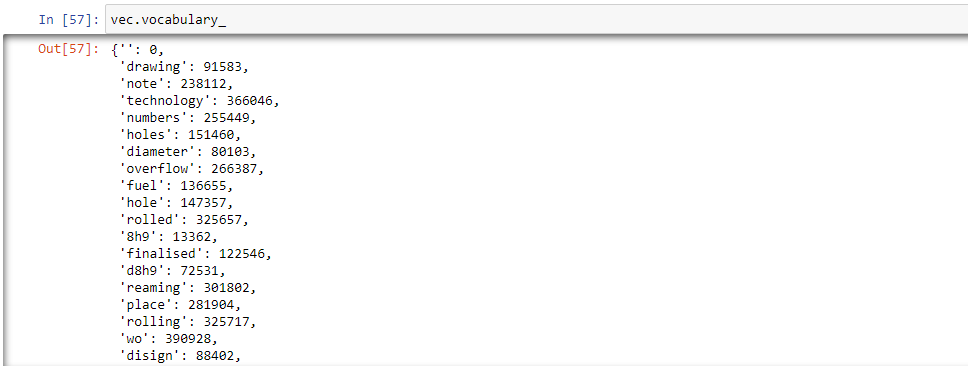


**Fig 4.3**

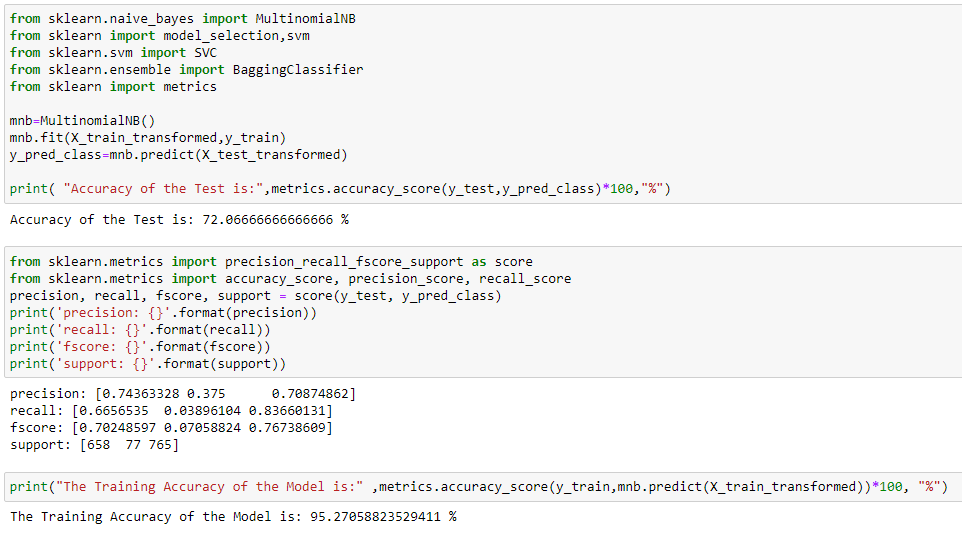
The fig 4.3 shows how the Dataset is divided into testing and training sets



**Fig 4.4**



With model selection done after initial testing model accuracy was 72.06%.

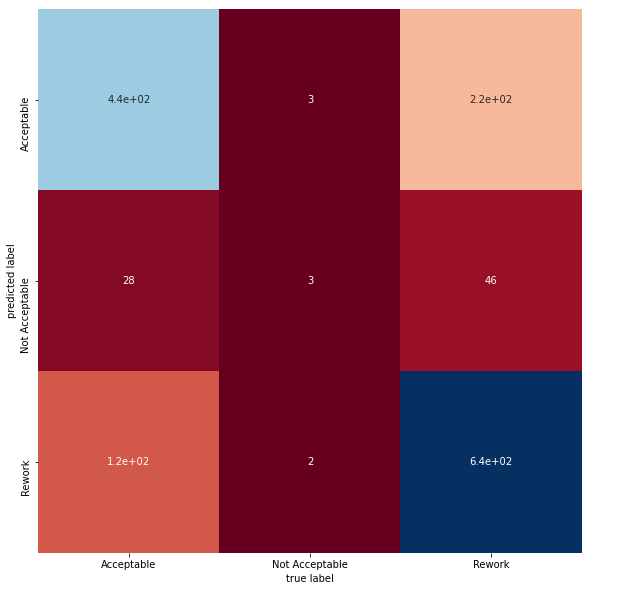


**Fig 4.5**



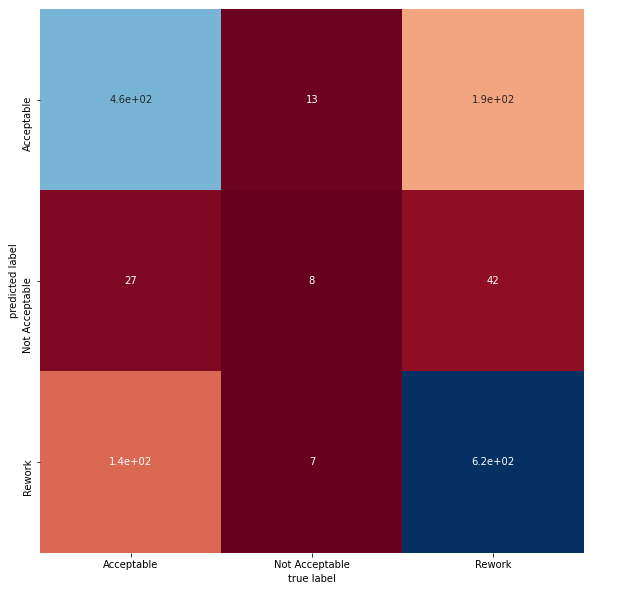
**Fig 4.6**

## Evaluation

Various parameters that are supported added to precisely tune model for our cause.

Though we tuned the model with maximum precision and fine values the accuracy scores were not satisfactory with only 76.05% accuracy with MultinomialNB

**Fig 4.7**

The accuracy score of SVM model is on average 83.01%. Heatmap of confusion matrix also shows how the majority classes got misclassified. With search performance model can’t be used for predicting real world snags .

**Fig 4.8**





## Model Fine Tuning

With such low accuracy, analysis had to be performed again to find the root cause of such behavior.

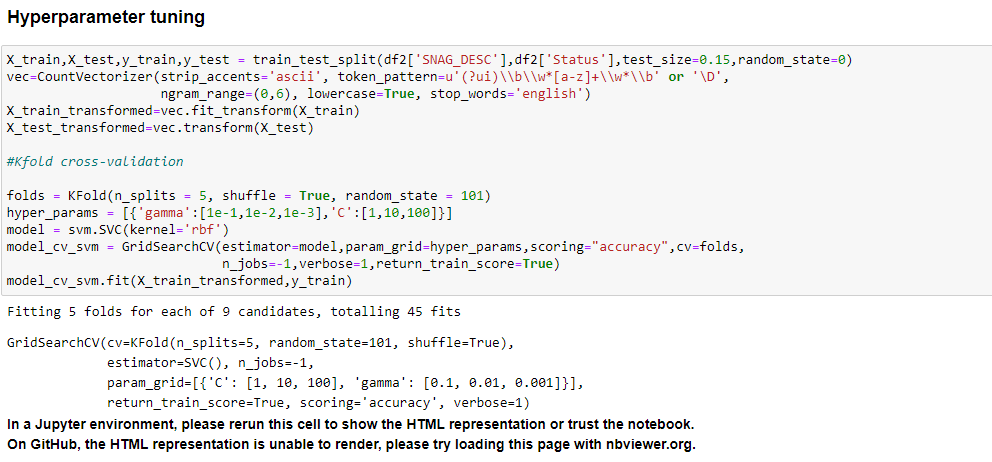
One of the problems were found to be spelling mistakes and different representation of the same word. This was fixed by manually changing the spellings as mentioned earlier .

Another problem which was a major one was found to be class imbalances.

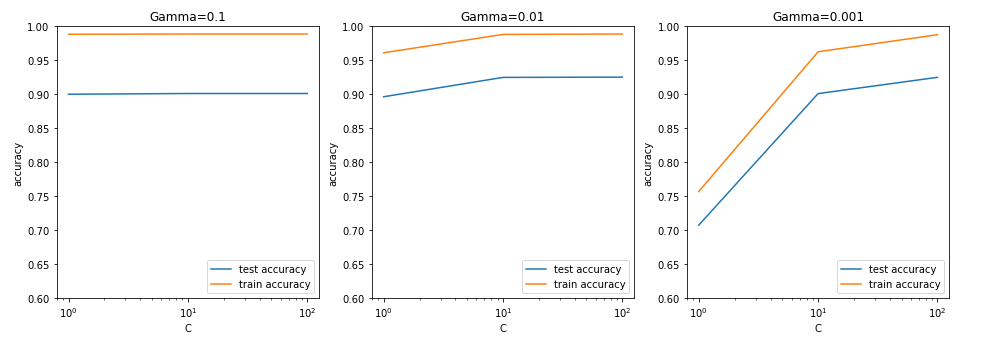
|  |  |
| --- | --- |
| **Fig 4.12** | **Fig 4.13** |

The target parameter Disposition had misproportions in the frequency of each type of class. This disproportionate class distribution needed to be fixed. Since the class “Not Acceptable” seems to be significantly less in frequency, we need to balance this class.

Data Upsampling algorithms were used to randomly create duplicates of data for the “Not Acceptable” class. This increased frequency of the “Not Acceptable” class. And fixed the issues of the irregular distribution of classes. The Fig 4.12 - 4.13 shows the class distribution before and after the class balancing.



**Fig 4.16**



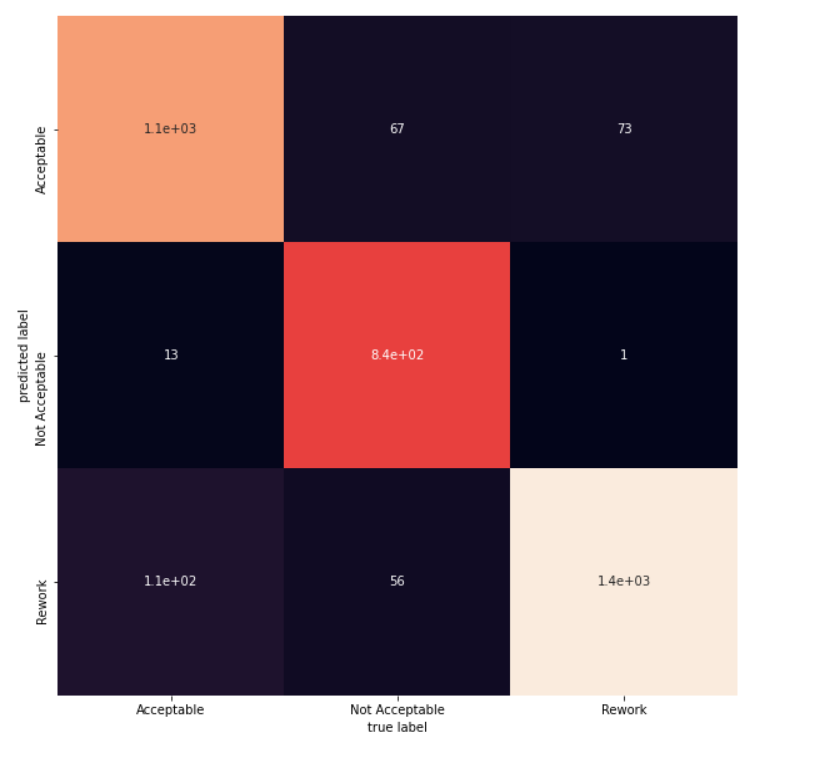
## Final Model Results

With updation done in dataset. The model was tuned for to accommodate the new changes. Once tuning was done, model was executed on the new dataset.

Under mentioned are the results :-



**Fig 4.16**



**Fig 4.16**

For SVM :

|  |
| --- |
| **Fig 4.16**  **The Accuracy of SVM model is 95.56 % and of Multinomial Naïve Bayes model is 91.23%.** |
|  |

As depicted in the figures the Multinomial Naïve Bayes model has a Testing Accuracy of **91.2347% and a Training Accuracy of 95.11%**. Furthermore **89.89%, 87.25%, and 94.97% accuracy for Acceptable, Not Acceptable, and Rework required respectively**.

As depicted in the figures the SVM model has a Testing Accuracy of **95.5495% and a Training Accuracy of 95.54%**. Furthermore **94.32%, 97.13%, and 95.64% accuracy for Acceptable, Not Acceptable, and Rework required respectively**

This level of accuracy seems acceptable for general-purpose use. Hence the ML model is ready. The model can now be deployed in the production stage.

## Results

Initially, the model was validated for 5 Live SNAGS out of which **all 5 SNAGS were predicted correctly**. Later on, the model was tested on 8 live SNAGS out of which 7 were predicted correctly with the last one being an outlier of the class. Appropriate changes were made to the dataset and following which **all 8 SNAGS were predicted correctly**.

The model can be deployed to various shops, assembly lines, and flight hangers for working.

# **CONCLUSION**

The project titled “AUTOMATIC SNAG DISPOSITION AND MANAGEMENT SYSTEM USING M.L”has been completed within the stipulated time of the project. The objective of the project was to Design and Develop an Machine Learning Model for the prediction of Snag Disposition based on SQMS dataset for training and testing, with live Snag taken as inputs from the user for assessment of the model. The model was successfully developed and is ready to be put in real-world systems.

This internship has given me a new perspective on various topics regarding military aircrafts. It has given a lot of exposure to the industrial side of a corporation. Through this, I got acquainted with the workflows, procedures, and safety measures involved in the industry. The Training Department gave me a brief knowledge about HAL, Nasik. The AURDC division gave me an in-depth overview of aircraft designing and various avionic system integration in aircrafts.

# **REFERENCES**

To gather information about the company and fighter aircrafts-

<https://hal-india.co.in/>

<https://en.wikipedia.org/wiki/Hindustan_Aeronautics_Limited>

<https://en.wikipedia.org/wiki/Sukhoi_Su-30MKI>

Theoretical knowledge and understandings-

<https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/>

<https://scikit-learn.org/stable/modules/naive_bayes.html>

<https://www.kaggle.com/code/mehmetlaudatekman/text-classification-svm-explained/notebook>

Documentation of various python 3 libraries used in the process-

<https://docs.python.org/3/>

<https://pandas.pydata.org/docs/>

<https://numpy.org/doc/>

<https://scikit-learn.org/stable/>

<https://matplotlib.org/>

<https://seaborn.pydata.org/>

Image sources –

By Alan Wilson - <https://www.flickr.com/photos/ajw1970/20261840172/in/album-72157656754662035/>, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=42136090>

By U.S. Air Force photo/ Airman 1st Class Ryan Crane - <http://www.mountainhome.af.mil/shared/media/photodb/photos/080717-F-1142C-188.jpg>, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=4429614>

Public Domain, <https://commons.wikimedia.org/w/index.php?curid=172130>