IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

A Statistical Analysis on Asthmatic Patient

Choudhar Shital Balu

Statistics Department, Tuljaram Chaturchand College, Baramati (Pune), Maharashtra, India.

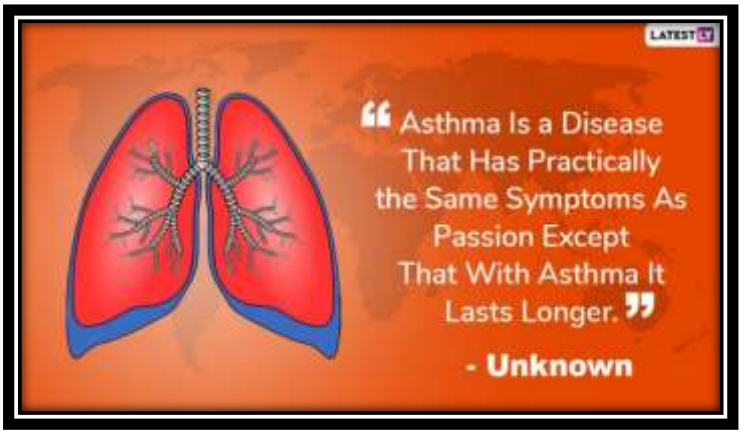
Abstract: 'Asthma' is derived from the Greek word **aazein** (ασθμαινω). **aazein** is 'short for breath'. The term originally did not define a disease, but was employed to describe respiratory symptoms of a variety of pulmonary conditions. This study includes analyzing the factors affecting on asthmatic patients and treatments of asthma.

Asthma was shown to be associated with transient increases in airway resistance, reductions in forced expiratory volumes and flows, hyperinflation of the lungs and increased work of breathing, as well as abnormalities in the distribution of ventilation, perfusion and arterial blood gases. Today, asthma is seen as a chronic inflammatory disease which is not yet fully understood in its pathophysiology; therefore, therapy is still on the path to becoming optimal.

Asthma is a common chronic disease in children. Uncontrolled asthma is a significant contributor to school absenteeism, emergency room visits, and hospitalization, all of which can lead to low school performance, financial burdens, and emotional problems for children and their parents.

Keywords: Asthma, symptoms, causes, impact, treatment

I. INTRODUCTION



Asthma is a chronic condition that affects your lungs. It's chronic condition, meaning it doesn't go away and needs ongoing medical management. It is also called "*Bronchial Asthma*". It is long term condition & it affects on the Air passage in Respiratory system.

It is observed in all age group but nowadays this chronic disease observed in adults. In current year sales of Asthmatic medicines is 43% in India. In India 42% deaths due to Asthma.

Symptoms of Asthma:

- Coughing
- Shortness of Breath / Rapid Breathing
- Wheezing
- Allergy
- Respiratory Infection
- Low o_2- level

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Causes of Asthma:

The most common asthma triggers include allergies, air pollution and other health conditions including respiratory infections, exercise or physical activity, weather and air temperature, strong emotions, and some medicines. Asthma is a chronic disease of the air passages characterized by inflammation and narrowing of the airways. Asthma, a major global health problem affecting as many as 235 million people worldwide. Symptoms of asthma include shortness of breath, cough, and wheezing. It commonly presents in childhood and is usually associated with conditions such as eczema and hay fever. This activity outlines the evaluation and treatment of asthma and explains the role of the interprofessional team in managing patients with this condition.

Asthma is a very common childhood illness leading to multiple hospital admissions and increased healthcare costs. The key feature is airway hyper-responsiveness, which can be triggered by many factors. If not treated promptly, asthma has a high mortality. Asthma is associated with exposure to tobacco smoke and inhaled particulates and is thus more common in groups with these environmental exposures. Asthma prevalence is greater in extreme of ages due to airway responsiveness and lower levels of lung function.

Following characteristic findings on examination

- Expiratory flow less than 33%
- Oxygen Level less than 92%
- Chest pain

Asthma treatment is based on a stepwise and control-based approach that involves an iterative cycle of assessment, adjustment of the treatment and review of the response aimed to minimize symptom burden and risk of exacerbations. Anti-inflammatory treatment is the mainstay of asthma management. In this review we will discuss the rationale and barriers to the treatment of asthma that may result in poor outcomes. The benefits of currently available treatments and the possible strategies to overcome the barriers that limit the achievement of asthma control in real-life conditions.

- **DATA COLLECTION**

The data relevant to this project has been taken from the UCI Machine Learning Repository which contains the Asthma Patient Dataset

- **SOFTWARE**
  - › Python
  - › R studio
  - › Microsoft Excel
  - › Microsoft Word

OBJECTIVES

- To predict whether the patient have Asthma or not.
- To predict preferences on Asthmatic Patients.
- To find a high performance predictive model that classifies Asthma.
- Check whether which treatment is effective with respect to Age.
- To determine health issue in Asthmatic patients.

METHODOLOGY

In this project we shall learn about the logistic regression and various classifiers used in Machine Learning to predict Asthma. We shall also explain our proposed methodology to improve the accuracy. Five different methods were used in this project. The different methods are defined below. The output is the accuracy matrix of the machine learning models. Then, the model can be used in prediction,

1. Decision Tree:

Decision Tree is the most powerful and popular tool for classification and prediction. Decision tree is type of supervised machine learning where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decision or final outcomes. And the decision nodes are where the data is split.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Decision tree can become much more powerful when used as ensembles. Ensembles is clever way of combining decision trees to create more powerful model. These ensembles create the state of the Art Machine Learning Algorithms that can outperform neural network in some cases.

2. Random Forest:

Random forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification and regression problem. In Machine learning. It is based on concept of ensembles learning. Which is a process of combining multiple classifiers to solve a complex problems and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contain a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of the dataset". Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predict the final output.

Advantages:

- Random forest is capable of performing both classification and regression techniques.
- It is capable of handling large dataset with high dimensionality.
- It enhanced the accuracy of the model and prevents the overfitting issue.

Disadvantages:

- Although random forest can be used for both classification and regression technique, it is not more suitable for

regression technique.

3. Logistic Regression model

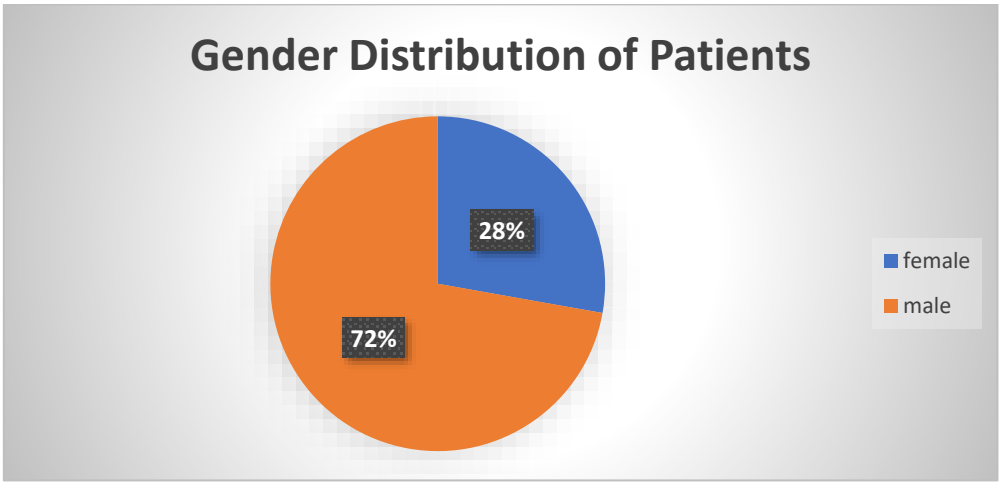
The logistic regression statistics modelling technique is used when we have a binary outcome variable.
Logistic regression is a statistical method that is used for building machine learning models where the dependent variable is dichotomous: i.e. binary. Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables. The independent variables can be nominal, ordinal, or of interval type.

The name "logistic regression" is derived from the concept of the logistic function that it uses. The logistic function is also known as the sigmoid function. The value of this logistic function lies between zero and one.

ASSUMPTIONS

- In logistic regression, dependent variable must be binary.
- For logistic regression, the factor level one of the dependent variables should represent the desired outcome
- The independent variables should be independent of each other. This means the model should have little or no multicollinearity
- Logistic regression requires quite large sample sizes.

Exploratory Data Analysis:



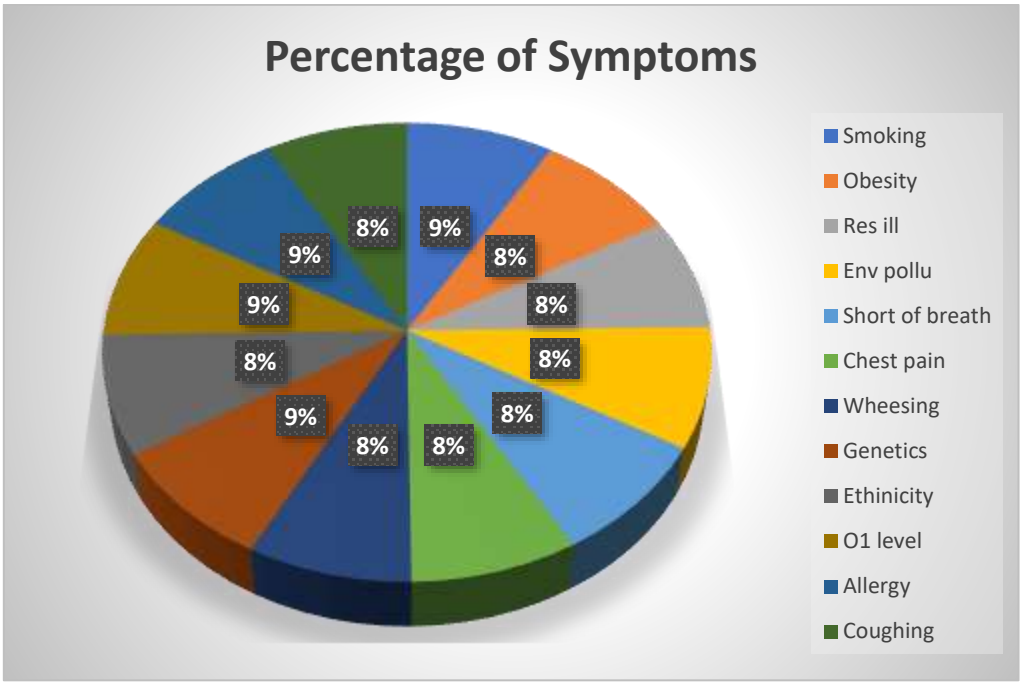
Conclusion: From above plot Number of Male Asthmatic Patients is 72% and Female Asthmatic Patients is 28%.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

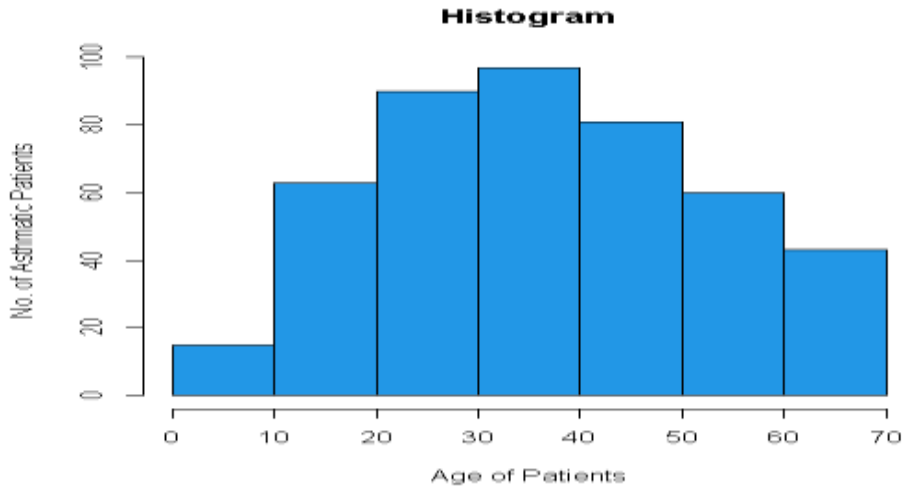
International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129



Conclusion: Above graph shows percentage of symptoms observed in asthma



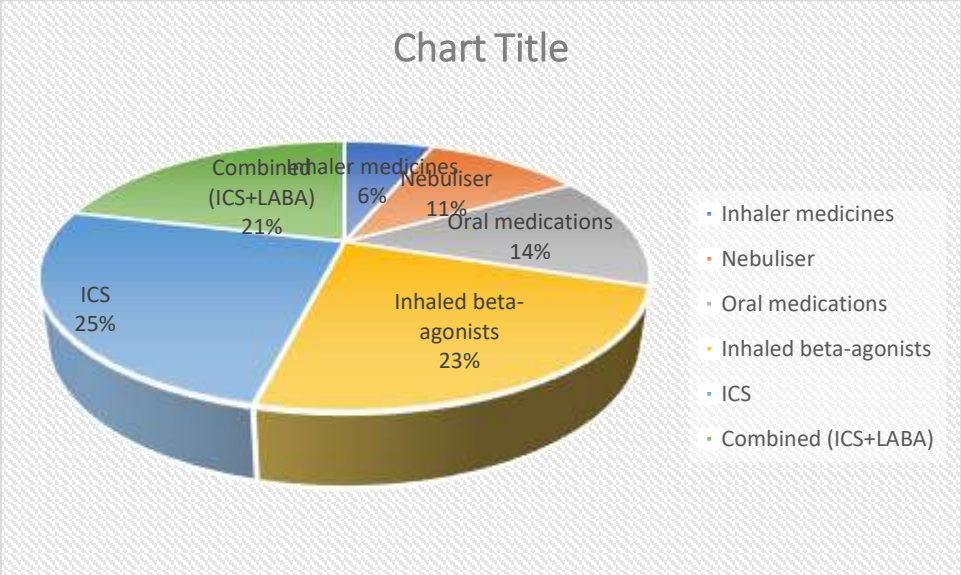
Conclusion: Above plot shows number of Asthmatic Patients corresponding to Age. From above plot we can say that Number of Asthmatic Patients are more in agegroup 20-50 i.e. more Youths are suffering from Asthma.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129



Conclusion: Above graph shows Inhaled Beta agonists, ICS and Combined (ICS+LABA) treatments are preferred by Asthmatic patients.

STATISTICAL ANALYSIS

1. Logistic Regression:

Dependent Variable*:* Y (A certain person is asthmatic or not)

Independent Variables*:* Age, Gender, Smoking, Obesity, Respiratory illness, Environmental pollution, Shortness of breath, Chest pain, Wheesing, Genetics, Ethnicity, O^2level, Allergy, coughing

Coefficients:

(Intercept)	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
1.999856	0.0056	-2.1313	1.02039	-0.8030	.20340	-0.53097	3.52326
X8	**X9**	**X10**	**X11**	**X12**	**X13**	**X14**	
2.902516	2.075610	-0.0842	-0.5449	1.1778	-2.0574	0.435309	

Null deviance: **557.90** on 499 degrees of freedom
Residual deviance: **404.91** on 485 degrees of freedom

Summary (Model):

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.49175	0.04167	0.52926	0.65322	1.90752

Coefficients:

	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	1.999856	0.505085	3.959	7.51e-05
Age	-0.005611	0.006817	-0.823	0.041045*
Gender	-2.131346	0.432113	-4.932	8.12e-07
smoking	1.020391	0.328084	3.110	0.018701*
obesity	-0.803065	0.496327	-1.618	0.105659
Res illness	-0.203448	0.495219	-0.411	0.038120*
Env pollution	0.530975	0.544038	0.976	0.329070

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Shortness of breath	-3.523261	0.708174	-4.975	0.025520*
Chest pain	2.902516	0.635847	4.565	0.001206*
wheesing	2.075610	0.522440	3.973	0.007100*
Genetics	-0.084263	0.539352	-0.156	0.875852
Ethnicity	-0.544967	0.511942	-1.065	0.287098
O2- level	-1.177883	0.507994	-2.319	0.020412*
Allergy	2.057346	0.622792	3.303	0.095532
Coughing	-0.435309	0.552285	-0.788	0.043058*

Model equation:
Y=1.999856+0.0056*Age-2.1313*Gender+1.02039*smoke-0.8030*obe+0.20340*Res ill-0.53097*Env poll+3.52326*short_breath+2.902516*chest_pain+2.075610*wheesing-0.0842*Gene-0.5449*Ethi+1.1778*o2_level-2.0574*Aller+0.435309*Cough

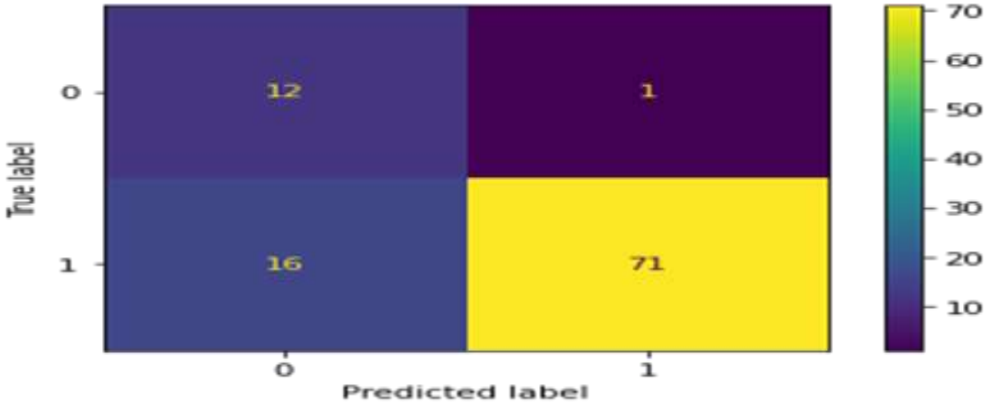
Conclusion: This model shows that prediction such as Age, Smoking, Respiratory illness, Shortage of breath, chest pain, whessing, Environmental pollution,o2-level and coughing this variables are significantly effect on asthma.

Deviance testing:
H0: Fitted model is Adequate v/s
H1: Fitted model is not Adequate
Decision criteria: Model Deviance > Chisq(α,n-p) then reject H0.
Model Deviance=152.99
Here α=0.05 ,p= 14 and n=500
Chisq(α,n-p)= 447.2509
Thus we failed to reject Ho
Conclusion: Fitted model is Adeuqate. i.e Model provides sufficient information about data.

Classification Report:

	precision	recall	f1-score	support
FALSE	0.43	0.92	0.59	13
TRUE	0.99	0.82	0.89	87
accuracy			0.89	100
macro avg	0.71	0.87	0.74	00
weighted avg	0.91	0.83	0.85	100

Confusion Matrix:



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Accuracy	83%
Precision	0.923077
sensitivity	0.428571
Specificity	0.986111
f1-score	89%

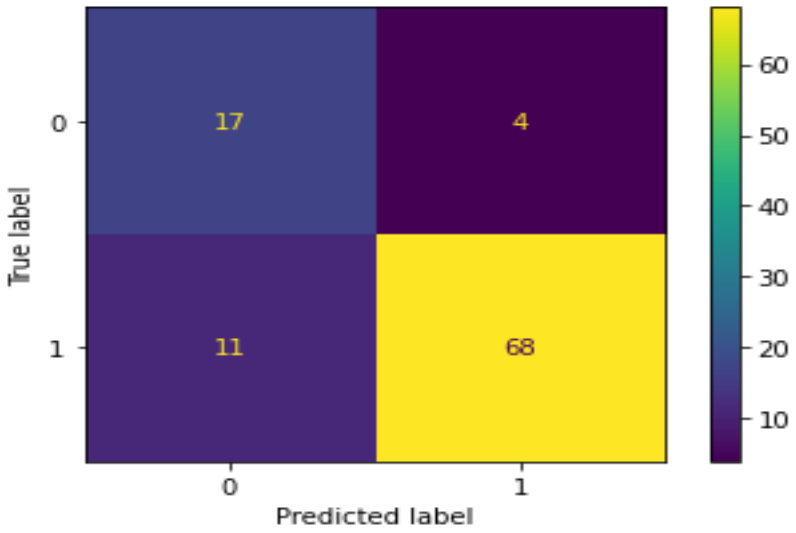
Conclusion: Logistic model gives 83% accuracy.

3. Decision Tree:

Classification Report:

	precision	recall	f1-score	support
FALSE	0.54	0.71	0.61	21
TRUE	0.92	0.84	0.87	79
accuracy		0.81	0.81	100
macro avg	0.73	0.77	0.74	100
weighted avg	0.84	0.81	0.82	100

Confusion Matrix:



Accuracy	83%
Precision	0.809524
sensitivity	0.607143
Specificity	0.944444
f1-score	89%

Conclusion: Decision Tree gives 85% accuracy.

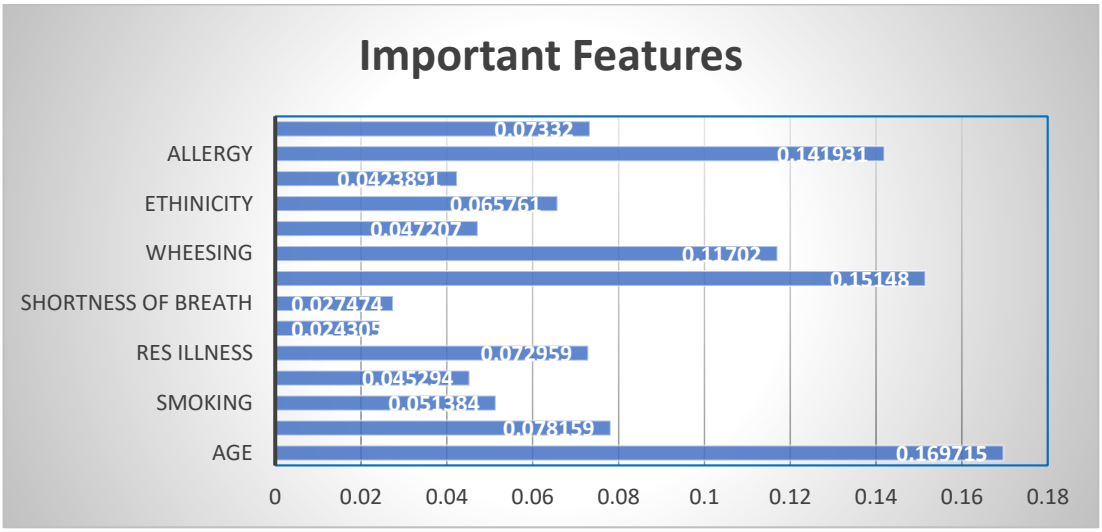
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Variables	Important Features
Age	0.169715
Gender	0.078159
Smoking	0.051384
Obesity	0.045294
Res illness	0.072959
Env pollution	0.024305
Shortness of breath	0.027474
Chest pain	0.15148
Wheesing	0.11702
Genetics	0.047207
Ethinicity	0.065761
O2- level	0.0423891
Allergy	0.141931
Coughing	0.07332



Conclusion: From table we get Age, Chest pain, Wheesing, Allergy are important variables.

3.Random Forest:

Classification Report:

	precision	recall	f1-score	support
FALSE	0.50	0.78	0.61	18
TRUE	0.94	0.83	0.88	82
accuracy			0.82	100
macro avg	0.72	0.80	0.75	100
weighted avg	0.86	0.82	0.83	100

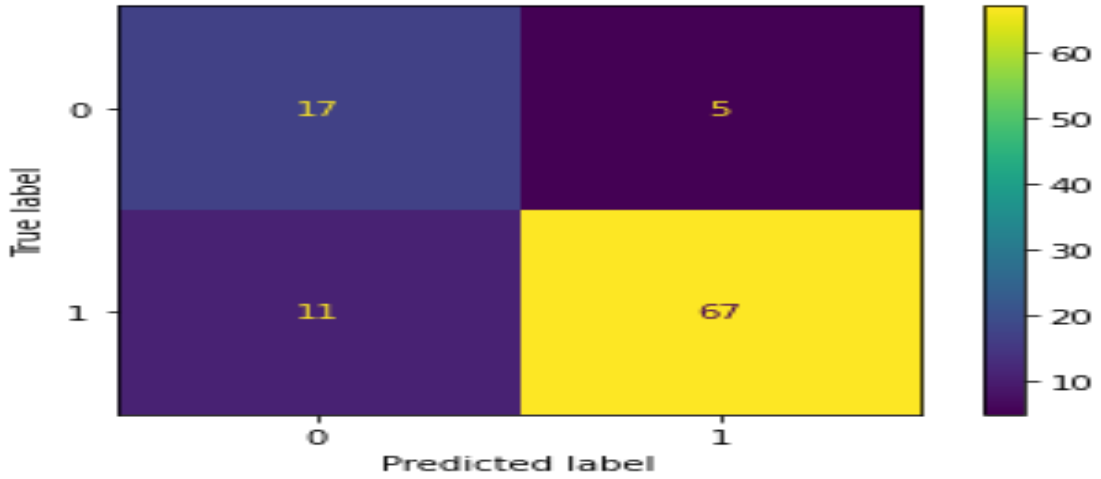
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

Confusion Matrix:



Accuracy	83%
Precision	0.772727
sensitivity	0.392857
Specificity	0.930556
f1-score	89%

Conclusion: Random forest gives 82% accuracy.

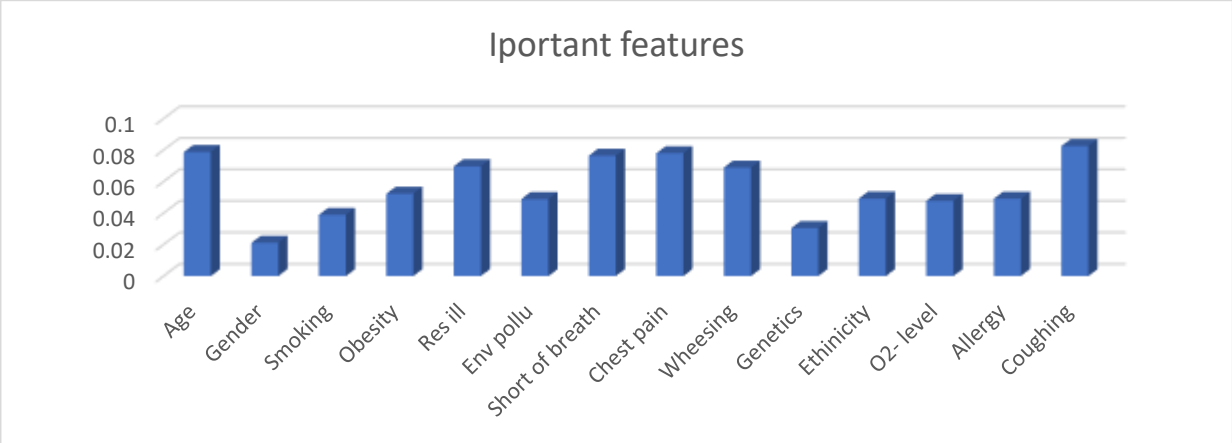
Variables	Iportant features
Age	0.0877956
Gender	0.04 15844
Smoking	0.058906
Obesity	0.052343
Res ill	0.06973
Env pollu	0.048896
Short of breath	0.0764223
Chest pain	0.077983
Wheesing	0.0688361
Genetics	0.03049
Ethinicity	0.049205
O2- level	0.04772836
Allergy	0.049089
Coughing	0.082534

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

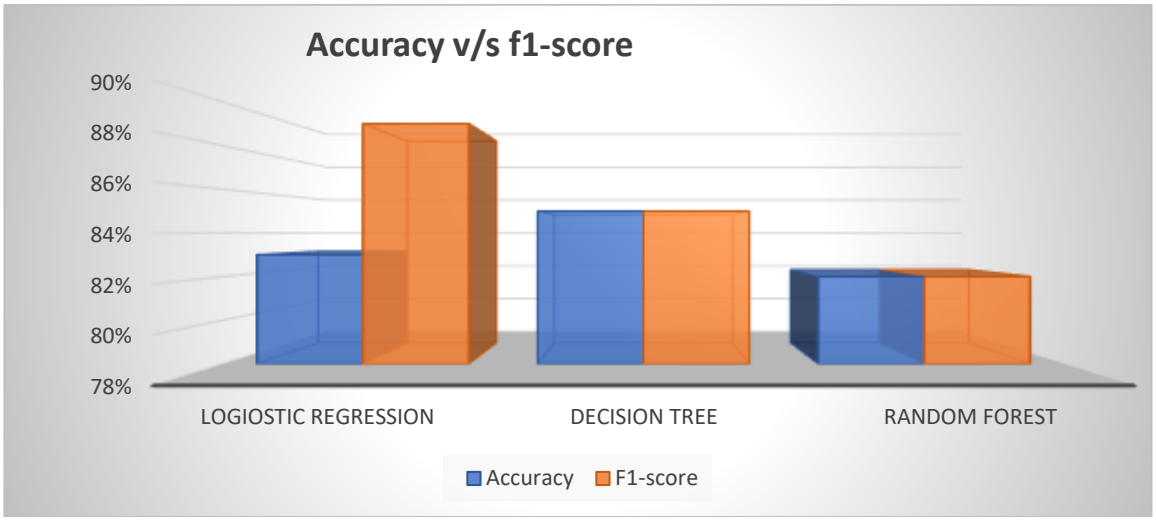
DOI: 10.17148/IARJSET.2025.12129



Conclusion: From table we get Age, Res illness, shortness of breath, Chest pain, Wheesing and coughing are important variables

CONCLUDING REMARKS

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score
Logiostic Regression	83%	0.923077	0.428571	0.986111	89%
Decision Tree	85%	0.779524	0.607143	0.944444	85%
Random forest	82%	0.912727	0.392857143	0.9305556	82%



Conclusion: From the above table we observe that accuracy of Decision tree is higher and accuracy of Logistic regression and Random forest is nearly equal. Due to high performance of Specificity and Sensitivity Logistic regression and Random forest are the best models than Decision tree for prediction of preference for Asthma.

CONCLUSION

- From exploratory data analysis we observe that number of male and female asthmatics.
- Inhaled Beta agonists, ICS and Combined (ICS+LABA) treatments are mostly preferred by Asthmatics.
- From logistic regression we observed that Age, Smoking, Respiratory illness, Shortage of breath, chest pain, whessing, Environmental pollution, o_2-level and coughing this variables are associate on asthma.
- From statistical models we observe that the Random forest and Logistic regression shows more accuracy.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 1, January 2025

DOI: 10.17148/IARJSET.2025.12129

REFERENCES

[1]. Official website of Agricultural marketing datahttps://agmarknet.gov.in

[2]. "Introduction to linear regression analysis" (Third Edition) Douglas C.Montegomery,Elizabeth A Peck, G Jeoffery, Vining

[3]. "Data mining concepts and techniques"(Third edition) Jiawei Han, Micheline kamber, Jain Pei

[4]. "Applied regression analysis"(Third edition) Norman R draper, Harry smith

[5]. Asthma Research and Practice 2023

[6]. The link between atopic dermatitis and asthma- immunological imbalance