

Lung Cancer Detection

We present a solution for identifying lung cancer in patients using a binary classification model.

 by Praveen Karthik Arumugam

LinkedIn - www.linkedin.com/in/pk7779

Github Portfolio - https://github.com/praveenkarthika/data_science.git



Problem Statement and Type

We aim to detect lung cancer in patients using the given dataset of symptoms. This is a Binary Classification problem.





Data Collection

Data is collected from www.kaggle.com and preprocessed for further analysis.

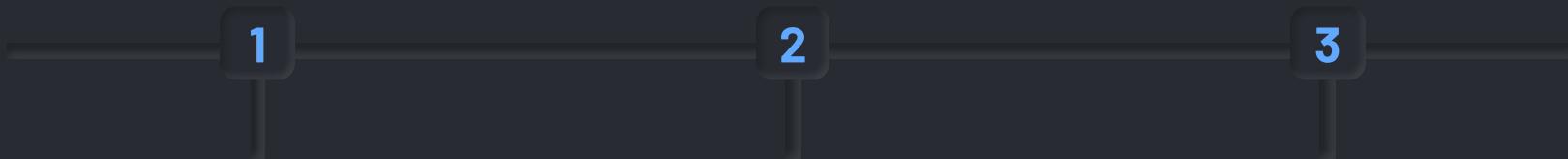
[Lung Cancer | Kaggle](https://www.kaggle.com/c/lung-cancer)

EDA - Exploratory Data Analysis

Feature Engineering

We identify the key features and extract useful information to enhance the accuracy of our model.

Model Development



Splitting dataset for Training and Testing

We split the dataset into training and testing sets in order to evaluate the performance of our model.

Feature Selection

We select the most relevant features to use in the model to obtain the highest accuracy.

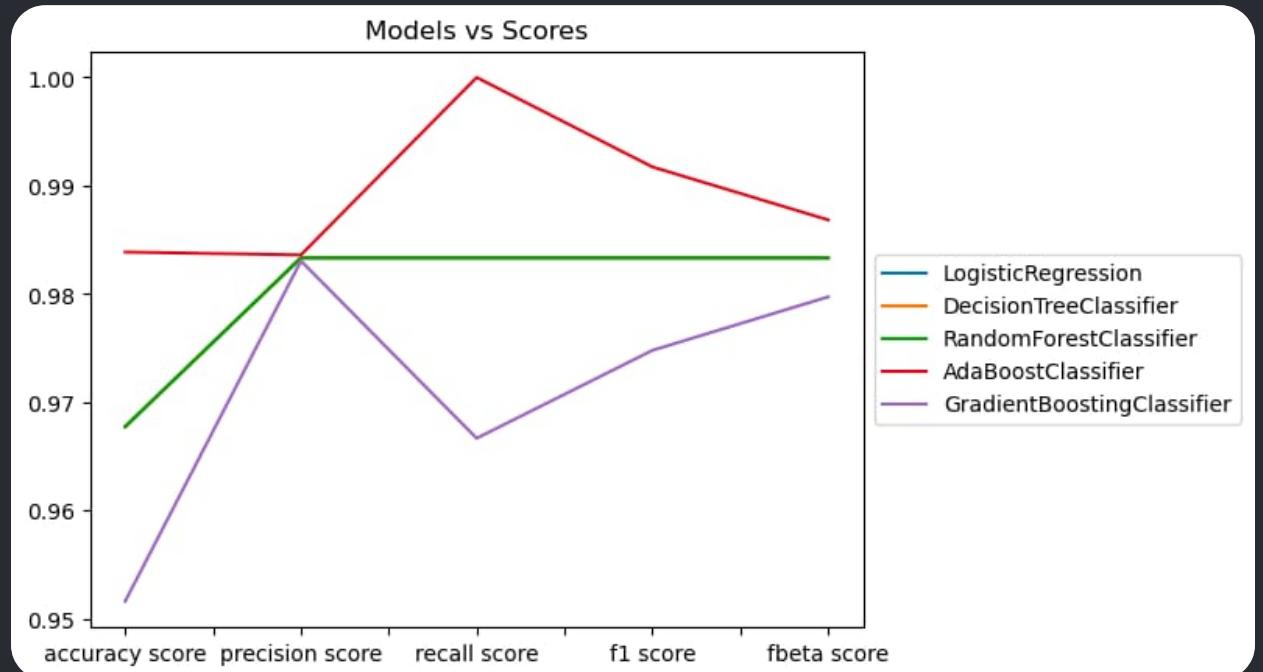
Model Selection

We use multiple models across Bagging and Boosting techniques and evaluate their performance using various accuracy metrics.

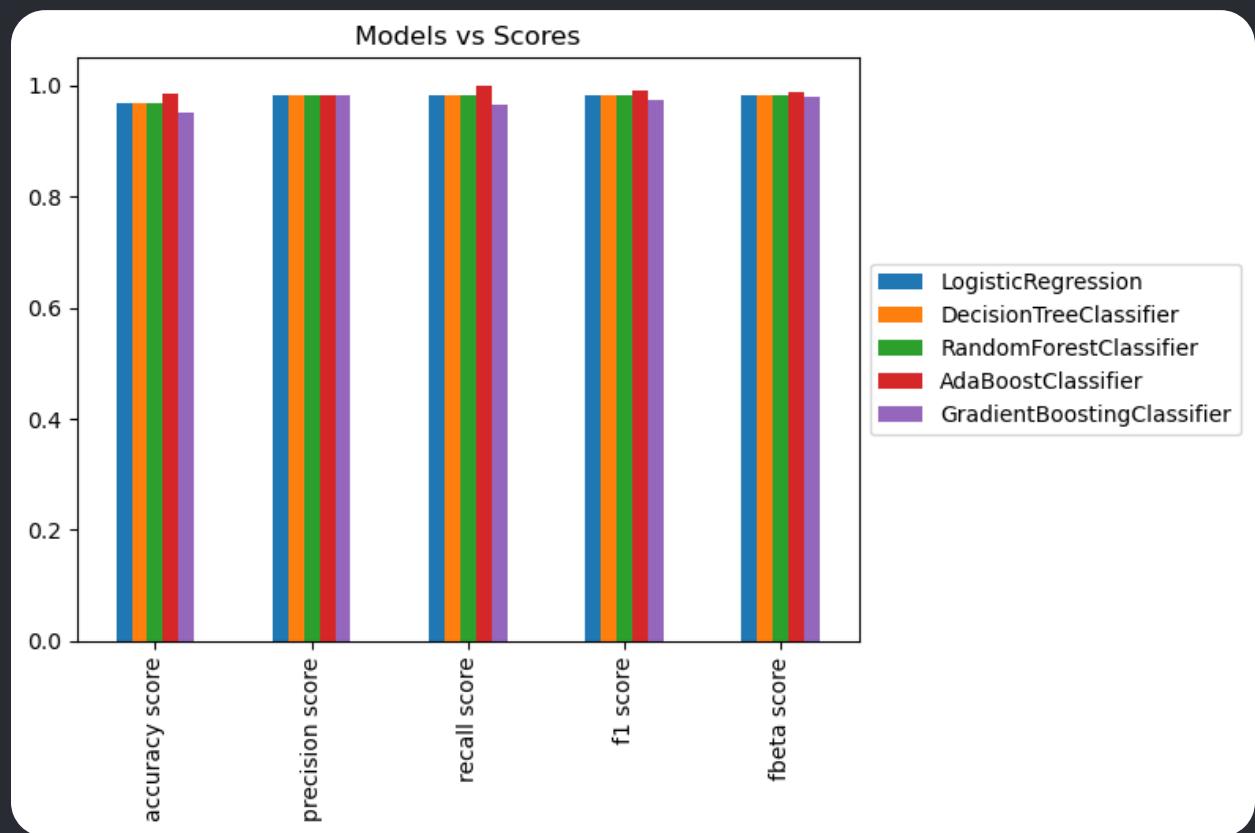
Model Performance Metrics

	Logistic Regression	Decision Tree Classifier	Random Forest Classifier	AdaBoost Classifier	Gradient Boosting Classifier
accuracy	0.967742	0.967742	0.967742	0.983871	0.951613
score					
precision	0.983333	0.983333	0.983333	0.983607	0.983051
score					
recall	0.983333	0.983333	0.983333	1.000000	0.966667
score					
f1 score	0.983333	0.983333	0.983333	0.991736	0.974790
fbeta	0.983333	0.983333	0.983333	0.986842	0.979730
score					

Models vs Scores - Line



Models vs Scores - Bar



About the Presenter

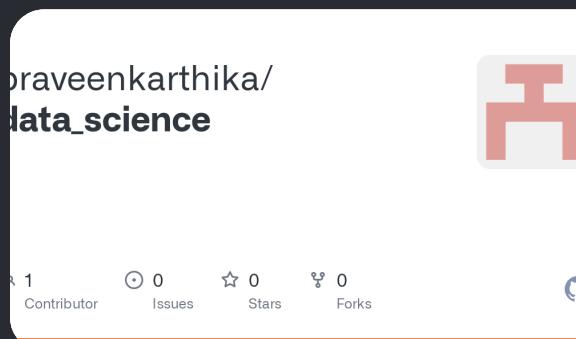
PraveenKarthik Arumugam

With a background in data science, Machine learning, Data Analytics and Automation/Manual Testing, PraveenKarthik Arumugam is an experienced professional in Information Technology.

Connect with him on LinkedIn at

www.linkedin.com/in/pk7779

Github Portfolio



GitHub

[GitHub - praveenkarthika/data_science](#)



Contribute to praveenkarthika/data_science development by creating an account on GitHub.