***Lung cancer survival Prediction***

**Name:** Rachita Rani

**Course:** BTech(CE)

**Name of Organization:**

Poornima University

Jaipur, Rajasthan

**Email Address:** [rachitaraj113@gmail.com](mailto:rachitaraj113@gmail.com)

Abstract— This e-document is prepared to define the components that have been used in the project and a description of the project

Keywords—seaborn,Matplotlib,algorithms,scilit-learn,sklearn,IEEE

# Introduction

This abstract document is being written to explain what does my project consists of and what are their basic components which had been used for its completion. This ML- project is basically prediction of the survival rate of the lung cancer patients. Through this we can be easily able to understand how we can predict all these with different categories. Various types of machine learning algorithms(ML) like Naive Bayes, Support Vector Machine (SVM), Logistic regression, Artificial Neural Network (ANN), have been applied in the healthcare sector for analysis of lung cancer.

## Abbreviations and Acronyms

ML(Machine Learning) is basically the branch of computer science that deals with algorithms and systems performing specific tasks using patterns and inference, rather than explicitly programmed instructions. IEEE( Institute of Electrical and Electronics Engineers).

## Libraries

* Pandas- **Pandas** is mainly **used** for data analysis. **Pandas** allows importing data from various file formats such as comma-separated values, JSON, SQL, Microsoft Excel.
* Numpy- **Numpy** provides a high-performance multidimensional array and basic tools to compute with and manipulate these arrays.
* Scikit-learn – It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy .
* Matplotlib- It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy .
* Seaborn - **Seaborn** is a library for making statistical graphics in Python.

# Creating This pRoject

This project has been created so that we can get basic knowledge on how to make a prediction module in ML. This project consist of dataset which is taken from kaggle. This module can helps us to analyze the lung cancer patients data and predict their survivality rate so that they can be treated accordingly.

## Authors and Affiliations

*This documentation is created by -*

*Rachita Rani*

## About

About 1 in 3 people with the condition **live** for at least 1 year after they're diagnosed and about 1 in 20 people **live** at least 10 years. However, survival rates vary widely, depending on how far the **cancer** has spread at the time of diagnosis. And day-to-day this rate is getting worsen and with the advancement of technologies these rates can be predicted easily and due to this prediction patients are treated accordingly so that they can improve their survival rate. This “Lung cancer survival prediction “ Module uses various libraries as mentioned above and explains the various algorithms that are used in Machine Learning for analysis and prediction.

##### DATA FLOW DIAGRAM

Key Learning: In this review, factors that cause lung cancer and application of ML algorithms are discussed up to date and also draws special attention to their relative strengths and weaknesses.

# 2.Paper title: A Comparative study of Lung cancer detection using Machine learning algorithms

Authors name: Radhika P.R.; Rakhi A.S. Nair; Veena G.

Year published: 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)

Key Learning: The key objective of this paper is the early diagnosis of lung cancer by examining the performance of classification algorithms.

# 3.Paper Title: A comparative study of machine learning techniques for the improved prediction of NSCLC survival analysis

Authors name: Alanna Vial; David Stirling; Matthew Field; Montserrat Ros; Christian Ritz; Martin Carolan; Lois Holloway; Alexis A. Miller

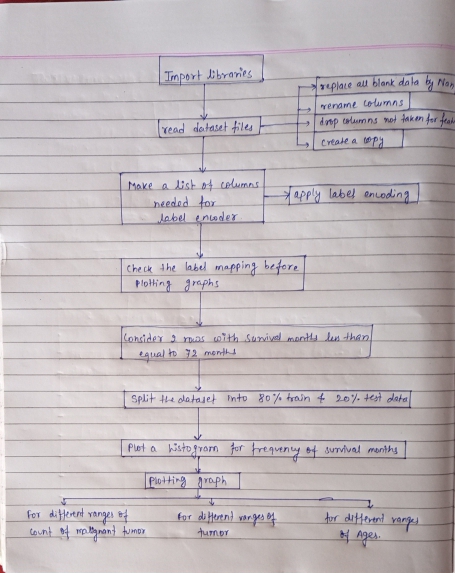
Year Published: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)

Key Learning: This paper challenges the traditional clinical ideas of radiotherapy where the centre of the tumor is treated with the highest dose, however this research indicates the periphery of the tumor is highly predictive of survival.

# 4.Paper Title: Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques

Authors name: James A. Bartholomai; Hermann B. Frieboes

Year Published: [2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)](https://ieeexplore.ieee.org/xpl/conhome/8637865/proceeding)

Fig: Flow chart

##### Acknowledgment

I would like to thank ExpertsHub for giving me this opportunity and Mr. Pavan Tej for his expert advice and encouragement throughout this difficult project, and also his brilliance in the practical implementation which he taught.

Literature Review

# 1.Paper title: Lung Cancer Prediction using Machine Learning: A Comprehensive Approach

Author name: Syed Saba Raoof; M A. Jabbar; Syed Aley Fathima

Year published: 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)

Key Learning: A regression model is developed to predict survival time in months for lung cancer patients. Comparison plots of the results further indicate that the regression models perform better for shorter survival times than the RMSE values are able to reflect.

# 5.Paper Title: Lung cancer survival prediction from pathological images and genetic data — An integration study

Authors name: Xinliang Zhu; Jiawen Yao; Xin Luo; Guanghua Xiao; Yang Xie; Adi Gazdar; Junzhou Huang

Year Published: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)

Key Learning: In this paper, a framework is proposed for lung cancer survival prediction by integrating genetic data and pathological images.

##### References

1. https://ieeexplore.ieee.org/abstract/document/9074947 *(references)*

*publishedin:* 2020 2nd internatinal conference [*https://ieeexplore.ieee.org/xpl/conhome/9057352/proceeding*](https://ieeexplore.ieee.org/xpl/conhome/9057352/proceeding)

1. <https://www.google.com/search?q=pandas&oq=pandas&aqs=chrome..69i57j69i60l2.1711j0j7&sourceid=chrome&ie=UTF-8>
2. <https://www.google.com/search?q=seaborn&oq=seaborn&aqs=chrome..69i57j0l7.3172j0j7&sourceid=chrome&ie=UTF-8>
3. <https://www.google.com/search?q=matplotlib&oq=matplotlib&aqs=chrome..69i57j69i60l2.3386j0j7&sourceid=chrome&ie=UTF-8>