INDUSTRIAL CERTIFICATION

Articial Intelligence using Google TensorFlow and Azure AI Fundamentals

A REPORT

submitted by

Omprakash (21BCE1950)

in partial fulfilment for the award

of

B. Tech Computer Science and Engineering

School of Computer Science and Engineering

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

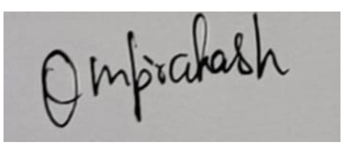
November 2024

School of Computer Science and Engineering

DECLARATION

I hereby declare that the project entitled **"Industrial Certification: "Articial Intelligence using Google TensorFlow" and Microsoft Certified: "Azure AI Fundamentals"** submitted by me to the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai 600127 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology – Computer Science and Engineering** is a record of bonafide work carried out by me**.** I further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or university.

Signature



Omprakash (21BCE1950)

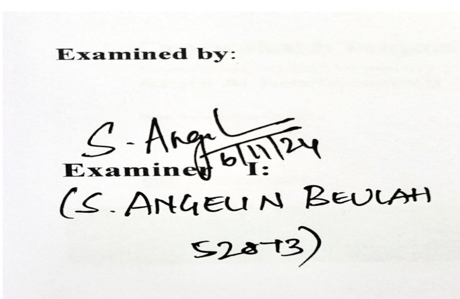
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

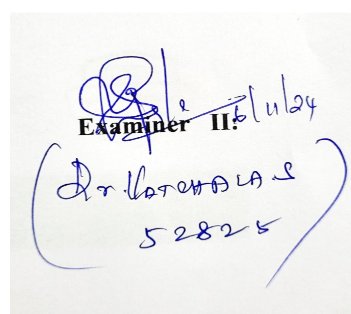
School of Computer Science and Engineering

CERTIFICATE

The project report entitled **"Industrial Certification: "Articial Intelligence using Google TensorFlow" and Microsoft Certified: "Azure AI Fundamentals"** is prepared and submitted by **OMPRAKASH (Register Number: 21BCE1950)**. It has been found satisfactory in terms of scope, quality and presentation as partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology – Computer Science and Engineering** in Vellore Institute of Technology, Chennai, India.

Examined by:


Examined by:
S. Angel 6/11/24
Examiner I:
(S. ANGELIN BEULAH
52873)


6/11/24
Examiner II:
(Dr. Vatchala S
52825)

Examiner I: **Examiner II:**

VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

CERTIFICATE

The project report entitled **"Industrial Certification: "Articial Intelligence using Google TensorFlow" and Microsoft Certified: "Azure AI Fundamentals"** is prepared and submitted by **OMPRAKASH (Register Number: 21BCE1950)**. It has been found satisfactory in terms of scope, quality and presentation as partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology – Computer Science and Engineering** in Vellore Institute of Technology, Chennai, India.

Examined by:

S. Angel 6/11/24

Examiner I:

(S. ANGELIN BEULAH
52873)

6/11/24

Examiner II:

(Dr Harshala S
52825)

3

Articial Intelligence using Google TensorFlow

SMARTBRIDGE
Let's Bridge the Gap
a Veranda Enterprise

In partnership with
India Edu Program
by
Google for Developers

Smart Internz

Certificate of Completion

This is to certify that Omprakash Pugazhendhi (21BCE1950) has successfully completed the Credit Course on Artificial Intelligence using Google TensorFlow Powered by Google Developers from June 2024 to July 2024 and fulfilled the project work requirements.

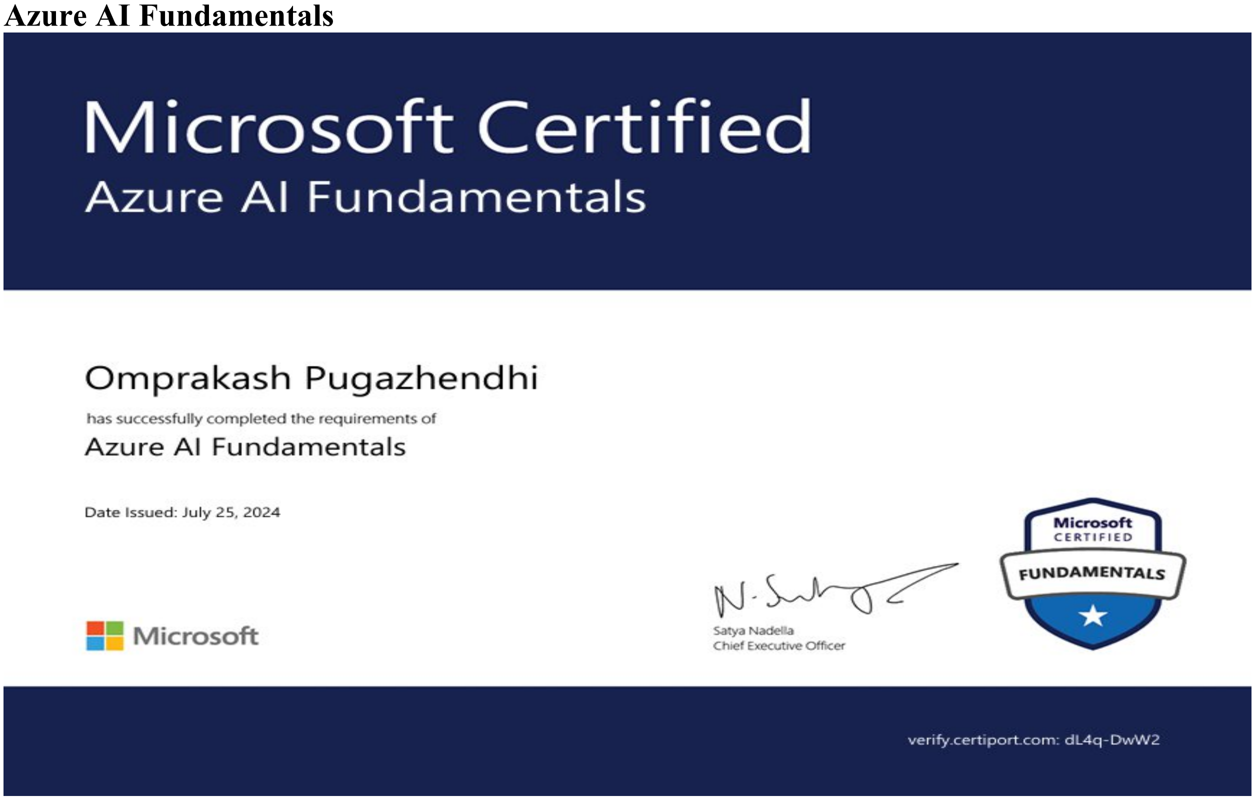
Certificate ID: CC-AI-2024-5248

August 12, 2024
Issued Date

SmartBridge Educational Services Pvt. Ltd. Hyderabad

Amarendar Katkam
Founder & CEO, SmartBridge

Azure AI Fundamentals



Microsoft Certified
Azure AI Fundamentals

Omprakash Pugazhendhi
has successfully completed the requirements of
Azure AI Fundamentals

Date Issued: July 25, 2024

Microsoft

Satya Nadella
Chief Executive Officer

Microsoft CERTIFIED FUNDAMENTALS

verify.certiport.com: dL4q-DwW2

ACKNOWLEDGEMENT

I would like to express my sincere gratitude for the successful completion of the **Articial Intelligence using Google TensorFlow and Azure AI Fundamentals** courses. The knowledge and experience gained from these two courses have been invaluable for my personal and professional growth.

I extend my heartfelt thanks to the Smart Bridge for providing the course which equipped me with the knowledge needed for the exam preparation. Additionally, I am grateful to the educational institutions that offered online courses for education and certification. I also express my gratitude to Microsoft for providing the opportunity to learn about azure cloud features. I would like to extend my thanks to Dr. Jayakumar Sadhasivam for providing all the course-related details and clearing all student queries. His expertise and support have been instrumental throughout the learning process

Furthermore, I would like to take this opportunity to express my deepest appreciation to:

Dr. Nithyanandam P, Head of the Department (HoD), B. Tech Computer Science and Engineering, SCSE, VIT Chennai,

Dr. Ganesan R, Dean of the School of Computer Science & Engineering, VIT Chennai,

Dr. Parvathi R, Associate Dean (Academics) of the School of Computer Science & Engineering, VIT Chennai,

Dr. Geetha S, Associate Dean (Research) of the School of Computer Science & Engineering, VIT Chennai, for their invaluable support, encouragement, and guidance throughout my academic journey. Their mentorship and expertise have been indispensable in shaping my research endeavours.

Place: Chennai — Omprakash

Date:06/11/2024 — Name of the student

CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Expansion
LSTM	Long Short-Term Memory
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
API	Application Programming Interface
EDA	Exploratory Data Analysis
ETL	Extract, Transform, Load
MLP	Multi-Layer Perceptron
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
ROC-AUC	Receiver Operating Characteristic - Area Under Curve
F1-Score	Model accuracy

ABSTRACT

In the rapidly evolving landscape of artificial intelligence (AI), the Artificial Intelligence using Google TensorFlow course emerges as a foundational stepping stone, equipping participants with an essential understanding of TensorFlow, one of the most widely used frameworks for machine learning and deep learning applications. This course addresses the growing demand for AI expertise in the current market, where businesses are increasingly leveraging AI and machine learning to gain insights, optimize operations, and create innovative solutions. However, the transition to AI-driven solutions is not without its challenges. Many organizations face hurdles such as a lack of skilled personnel, data quality issues, and the complexity of implementing and managing AI models effectively.

The Artificial Intelligence using Google TensorFlow course tackles these issues head-on by providing a thorough overview of AI concepts, TensorFlow's core tools and libraries, model development, training, evaluation, and deployment. By demystifying the complexities of TensorFlow, the course prepares individuals to design and implement AI solutions based on real-world business requirements. Furthermore, it lays a solid foundation for further specialization in AI and deep learning, addressing the industry's acute need for AI-savvy professionals. As such, this course is not just a pathway to certification but a crucial investment in a future-proof career in artificial intelligence and machine learning.

The Artificial Intelligence using Google TensorFlow course is a comprehensive training program that provides a thorough introduction to TensorFlow's powerful capabilities. The course covers essential topics, such as building neural networks, working with large datasets, optimizing model performance, and deploying AI models at scale. Through hands-on exercises and labs, participants gain practical experience working with TensorFlow, preparing them to effectively implement and manage AI models in production environments. Upon completion of the course, participants are equipped with the knowledge and skills needed to pursue advanced certifications in AI and machine learning, validating their expertise in the rapidly growing field of artificial intelligence.

In conclusion, the Artificial Intelligence using Google TensorFlow course plays a pivotal role in addressing the current limitations and challenges faced by businesses in AI and machine learning. By equipping participants with the necessary skills and knowledge, this course not only solves immediate problems but also prepares individuals and organizations for the future, ensuring they remain competitive in a rapidly changing digital landscape.

1. INTRODUCTION

In an era where digital transformation dictates the pace of business innovation, the necessity for professionals to possess specialized knowledge in artificial intelligence (AI) and cloud computing has never been more critical. This report delves into the significance of two pivotal certifications in the technology domain: Artificial Intelligence using Google TensorFlow and Azure AI Fundamentals. These certifications not only represent a benchmark of expertise for individuals in their respective fields but also highlight the evolving landscape of IT, where AI and cloud infrastructure are emerging as core pillars of modern business strategies.

The Artificial Intelligence using Google TensorFlow certification is designed as a comprehensive qualification for individuals looking to validate their proficiency in AI and machine learning. TensorFlow, an open-source framework developed by Google, has become one of the most popular platforms for building and deploying AI models. This certification covers essential AI concepts, including deep learning, neural networks, model training, and optimization using TensorFlow. It aims to equip professionals with the technical skills required to design, develop, and deploy AI solutions, making it an invaluable asset for anyone seeking to establish expertise in the rapidly growing field of AI.

On the other hand, the Azure AI Fundamentals course is a foundational training program that introduces participants to key AI concepts and the Azure AI ecosystem. Azure, Microsoft's cloud computing platform, offers a comprehensive suite of AI services, making it a popular choice for organizations looking to implement AI-driven solutions. The Azure AI Fundamentals course covers a range of topics, including machine learning, computer vision, natural language processing, and AI solutions in the cloud. It prepares participants with the necessary knowledge to start working with Azure AI services, helping businesses leverage AI for innovation and efficiency.

This report aims to explore the significance of these certifications, highlighting how they prepare individuals to tackle the challenges and opportunities presented by AI and cloud technologies. Through an examination of the current technology landscape, this analysis will shed light on how these certifications enable professionals to contribute to their organizations' digital transformation efforts. By providing a detailed overview of the content, benefits, and real-world applications of the Artificial Intelligence using Google TensorFlow and Azure AI Fundamentals courses, this report seeks to underscore the value of these certifications in fostering a future-ready workforce.

In conclusion, as we stand on the brink of a new digital epoch, the Artificial Intelligence using Google TensorFlow and Azure AI Fundamentals certifications emerge as key enablers of technological proficiency. This report explores how these certifications not only equip individuals with the requisite knowledge and skills but also serve as catalysts for innovation and growth in the dynamic world of AI and cloud computing.

2. AI USING GOOGLE TENSORFLOW

The coexistence of liver diseases poses significant clinical challenges, requiring effective predictive models for early detection and intervention. In this study, we employed decision tree and logistic regression algorithms to predict the likelihood of liver disease in individuals diagnosed . Distinct datasets were utilized, for liver disease prediction, containing relevant clinical attributes. Through rigorous experimentation and evaluation, our models demonstrated promising performance in identifying the presence of liver disease in individuals.

In the ever-evolving field of healthcare, predicting and preventing liver diseases have become paramount to ensuring the well-being of individuals and communities. Today, we will delve into two powerful machine learning techniques, Logistic Regression and Decision Tree, which have shown significant potential in predicting the likelihood of these diseases. Logistic Regression is a statistical method that allows us to model the relationship between predictor variables and a binary outcome, such as the presence or absence of liver diseases. This technique is particularly useful

when we want to understand the effect of various factors on the probability of a specific disease. Decision Trees, on the other hand, are a non- parametric method used for both classification and regression tasks. They work by recursively splitting the data into subsets based on the most significant predictor variables, thus creating a tree- like model that can be easily interpreted and understood. In the context of predicting liver diseases, decision trees can help identify the most important risk factors and provide a visual representation of the decision-making process. Combining these two techniques can lead to more accurate and robust predictions, as well as a deeper understanding of the complex interplay between various risk factors and the likelihood of developing liver diseases. By employing these machine learning algorithms, researchers and healthcare professionals can develop personalized preventive measures, early detection strategies, and more effective treatments to improve overall patient outcomes. In conclusion, the integration of Logistic Regression and Decision Tree in predicting liver diseases holds great promise for advancing healthcare and saving lives. As we explore these techniques further, we can expect to gain valuable insights into disease risk factors and contribute to the development of more effective, personalized healthcare strategies.

Google Tensorflow Overview

The coexistence of liver diseases poses significant clinical challenges, requiring effective predictive models for early detection and intervention. In this study, we employed decision tree and logistic regression algorithms to predict the likelihood of liver disease in individuals diagnosed . Distinct datasets were utilized, for liver disease prediction, containing relevant clinical attributes. Through rigorous experimentation and evaluation, our models demonstrated promising performance in identifying the presence of liver disease in individuals.

In the ever-evolving field of healthcare, predicting and preventing liver diseases have become paramount to ensuring the well-being of individuals and communities. Today, we will delve into two powerful machine learning techniques, Logistic Regression and Decision Tree, which have shown significant potential in predicting the likelihood of these diseases. Logistic Regression is a statistical method that allows us to model the relationship between predictor variables and a binary outcome, such as the presence or absence of liver diseases. This technique is particularly useful

when we want to understand the effect of various factors on the probability of a specific disease.

Decision Trees, on the other hand, are a non- parametric method used for both classification and regression tasks. They work by recursively splitting the data into subsets based on the most significant predictor variables, thus creating a tree- like model that can be easily interpreted and understood. In the context of predicting liver diseases, decision trees can help identify the most important risk factors and provide a visual representation of the decision-making process. Combining these two techniques can lead to more accurate and robust predictions, as well as a deeper understanding of the complex interplay between various risk factors and the likelihood of developing liver diseases. By employing these machine learning algorithms, researchers and healthcare professionals can develop personalized preventive measures, early detection strategies, and more effective treatments to improve overall patient outcomes. In conclusion, the integration of Logistic Regression and Decision Tree in predicting liver diseases holds great promise for advancing healthcare and saving lives. As we explore these techniques further, we can expect to gain valuable insights into disease risk factors and contribute to the development of more effective, personalized healthcare strategies.

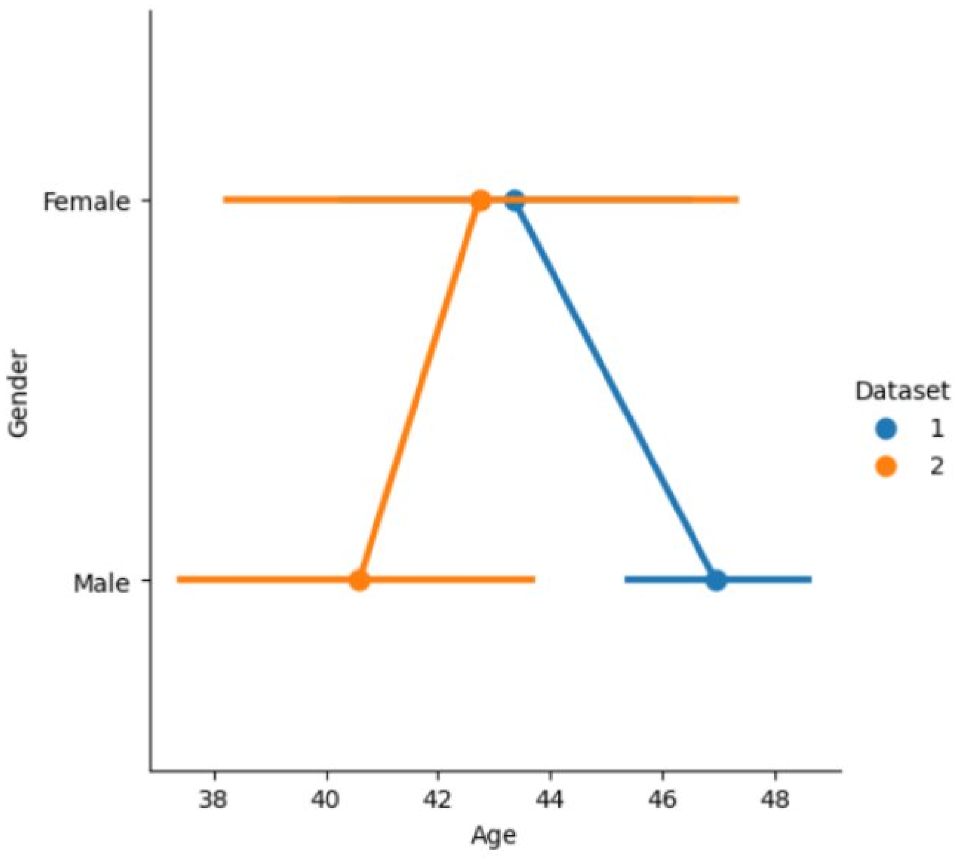
```
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    float64
 3   Direct_Bilirubin            583 non-null    float64
 4   Alkaline_Phosphotase        583 non-null    int64
 5   Alamine_Aminotransferase    583 non-null    int64
 6   Aspartate_Aminotransferase  583 non-null    int64
 7   Total_Protiens              583 non-null    float64
 8   Albumin                     583 non-null    float64
 9   Albumin_and_Globulin_Ratio  579 non-null    float64
 10  Dataset                     583 non-null    int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

indian-liver-patient	Outliers: The dataset contains some features with a wide range of values, indicating the potential presence of outliers that should be identified and handled appropriately.	Moderate	Some of the feature values, such as Alamine_Aminotransferase, Aspartate_Aminotransferase, and Total_Bilirubin, appear to have a wide range, indicating the potential presence of outliers. Outliers can significantly impact model performance and should be identified and handled appropriately.
indian-liver-patient	Data Quality: The dataset contains some duplicate rows, which should be identified and removed to ensure data integrity	Moderate	The dataset has a large number of rows (over 500) and a wide range of values for the different features, indicating it may be a comprehensive dataset. However, there are a few rows with duplicate values for some patients, which could indicate data quality issues that need to be addressed
indian-liver-patient	Imbalanced Classes: The target variable (liver disease diagnosis) is imbalanced, with more instances of liver disease (class 1) than no liver disease (class 2). This may require techniques like oversampling or undersampling to balance the classes	Low	The "Dataset" column indicates that the target variable (liver disease diagnosis) is imbalanced, with 1 representing liver disease and 2 representing no liver disease. Imbalanced target variables can pose challenges for machine learning models and may require techniques like oversampling or undersampling to address

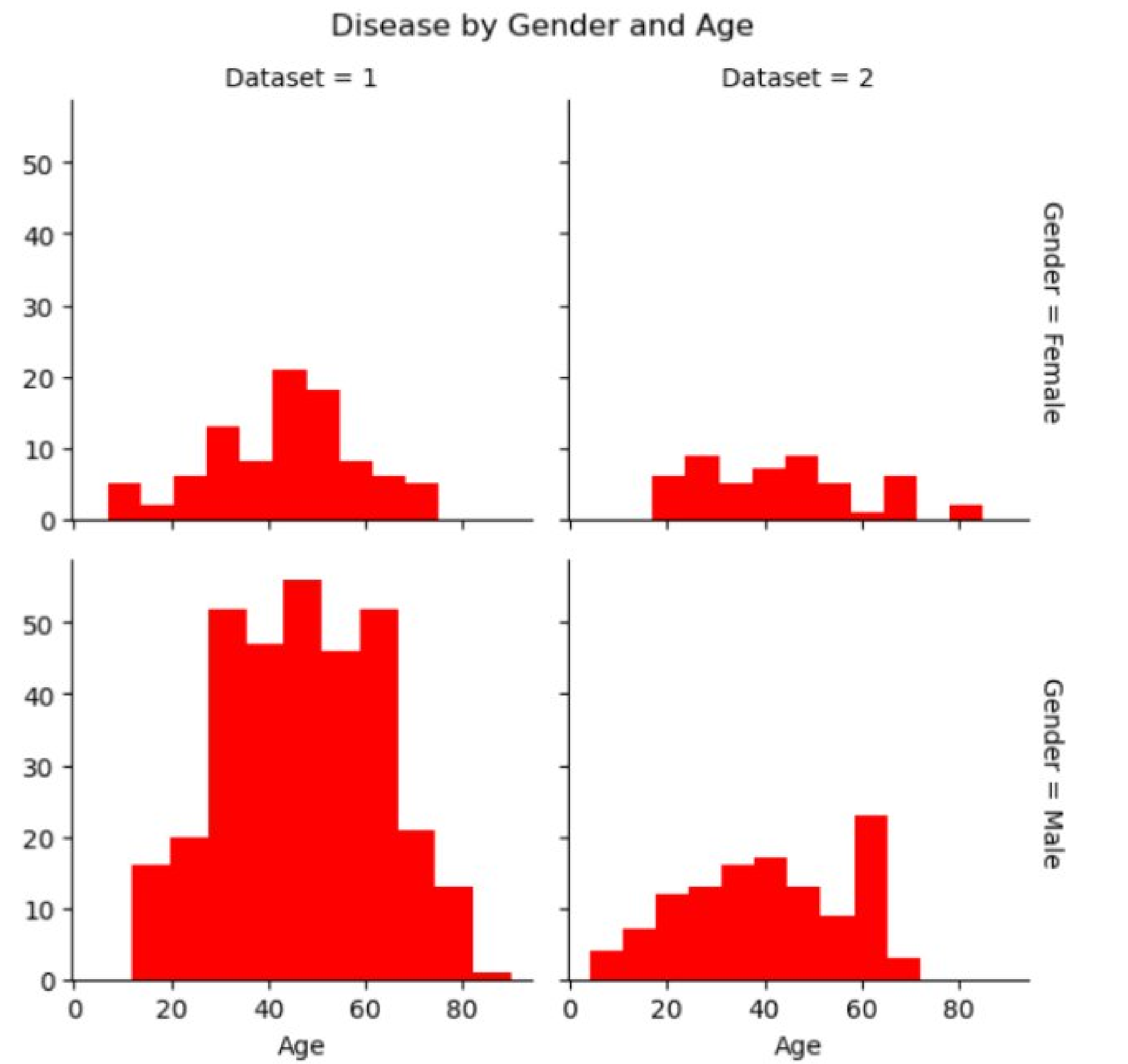
	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin
count	583.000000	583	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000
unique	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	NaN	Male	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	NaN	441	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	44.746141	NaN	3.298799	1.486106	290.576329	80.713551	109.910806	6.483190	3.141852
std	16.189833	NaN	6.209522	2.808498	242.937989	182.620356	288.918529	1.085451	0.795519
min	4.000000	NaN	0.400000	0.100000	63.000000	10.000000	10.000000	2.700000	0.900000
25%	33.000000	NaN	0.800000	0.200000	175.500000	23.000000	25.000000	5.800000	2.600000
50%	45.000000	NaN	1.000000	0.300000	208.000000	35.000000	42.000000	6.600000	3.100000
75%	58.000000	NaN	2.600000	1.300000	298.000000	60.500000	87.000000	7.200000	3.800000
max	90.000000	NaN	75.000000	19.700000	2110.000000	2000.000000	4929.000000	9.600000	5.500000

```
sns.catplot(x="Age", y="Gender", hue="Dataset", data=liver_df, kind="point")
```

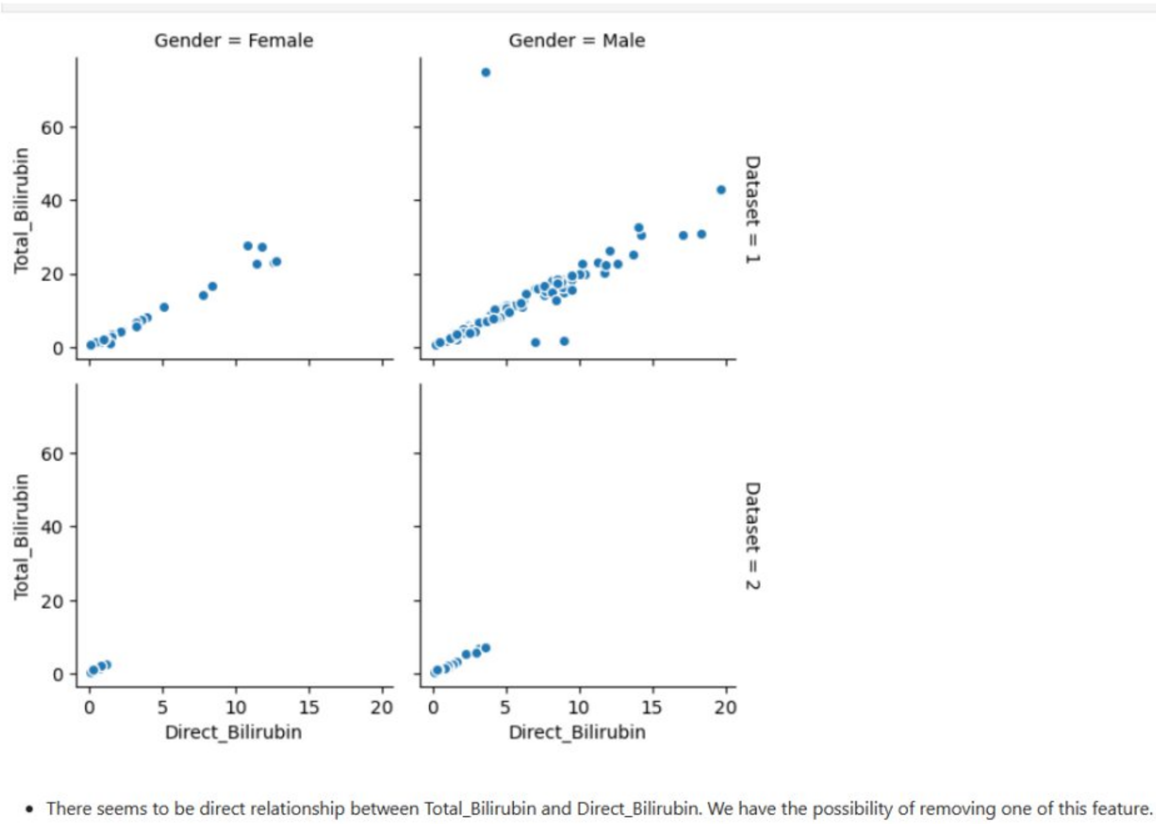
```
<seaborn.axisgrid.FacetGrid at 0x1b6818204d0>
```



- Age seems to be a factor for liver disease for both male and female genders

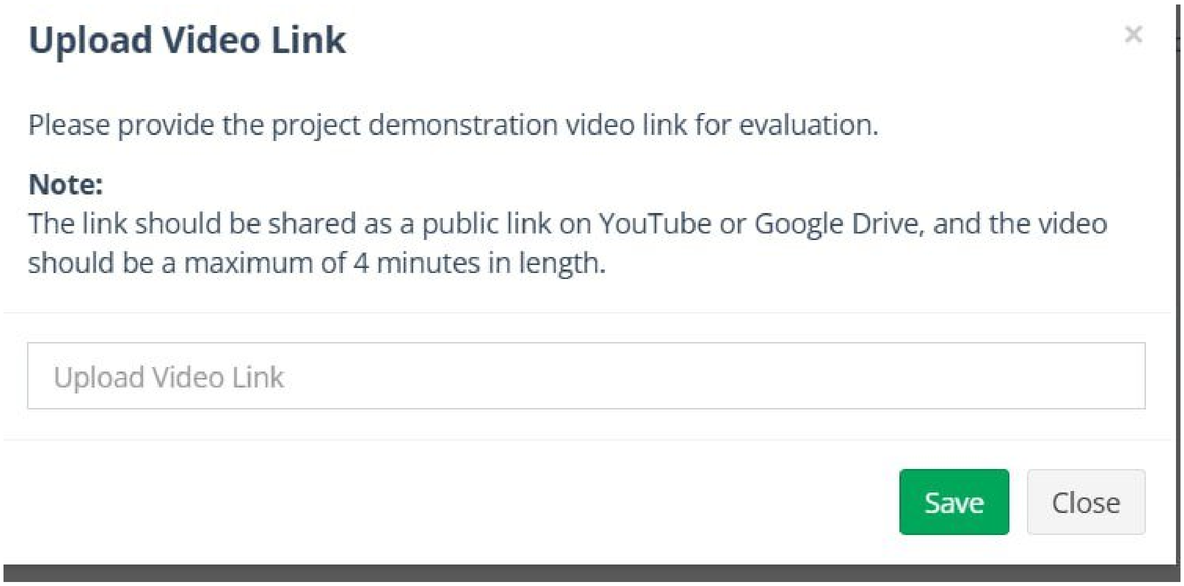


Disease by Gender and Age
Dataset = 1
Dataset = 2
Gender = Female
Gender = Male
0
10
20
30
40
50
0
20
40
60
80
Age
Age



The coexistence of liver diseases poses significant clinical challenges, requiring effective predictive models for early detection and intervention. In this study, we employed decision tree and logistic regression algorithms to predict the likelihood of liver disease in individuals diagnosed . Distinct datasets were utilized, for liver disease prediction, containing relevant clinical attributes. Through rigorous experimentation and evaluation, our models demonstrated promising performance in identifying the presence of liver disease in individuals.
In the ever-evolving field of healthcare, predicting and preventing liver diseases have become paramount to ensuring the well-being of individuals and communities. Today, we will delve into two powerful machine learning techniques, Logistic Regression and Decision Tree, which have shown significant potential in predicting the likelihood of these diseases. Logistic Regression is a statistical method that allows us to model the relationship between predictor variables and a binary outcome, such as the presence or absence of liver diseases. This technique is particularly useful when we want to understand the effect of various factors on the probability of a specific disease. Decision Trees, on the other hand, are a non-parametric method used for both classification and regression tasks. They work by recursively splitting the data into subsets based on the most significant predictor variables, thus creating a tree- like model that can be easily interpreted and understood. In the context of predicting liver diseases, decision trees can help identify the most important risk factors and provide a visual representation of the decision-making process. Combining these two techniques can lead to more accurate and robust predictions, as well as a deeper understanding of the complex interplay between various risk factors and the likelihood of developing liver diseases. By employing these machine learning algorithms, researchers and healthcare professionals can develop personalized preventive measures, early detection strategies,

and more effective treatments to improve overall patient outcomes. In conclusion, the integration of Logistic Regression and Decision Tree in predicting liver diseases holds great promise for advancing healthcare and saving lives. As we explore these techniques further, we can expect to gain valuable insights into disease risk factors and contribute to the development of more effective, personalized healthcare strategies.



KNN	AUC-ROC: 0.6477162293488825	accuracy 0.66
Random forest classifier	AUC-ROC: 0.7103624518590504	accuracy 0.68
Logistic regression	AUC-ROC: 0.6906921498758234	accuracy 0.67

Final Model	Reasoning
Logistic regression	Has high accuracy when compared with other two models High recall score and high precision score The model strikes a balance between interpretability and good results

Model Selection Report:

Model-1(Logistic regression)

Description:

One of the simplest and best ML classification algorithms is logistic regression. LR is a supervised ML binary classification algorithm widely used in most applications. It operates on a categorical dependent variable, the result can be a discrete or binary categorical variable 0 or 1.

Logistic sigmoid function:

$$prob(Y=1)=\frac{e^z}{1+e^z}$$

Hyperparameters:

```
{'C': 0.0001, 'max_iter': 1000, 'solver': 'lbfgs'}
▾ LogisticRegression
LogisticRegression(C=0.0001, max_iter=1000)
```

Accuracy, Precision and Recall:

	precision	recall	f1-score	support
0	0.44	0.46	0.45	147
1	0.78	0.77	0.78	378
accuracy			0.68	525
macro avg	0.61	0.61	0.61	525
weighted avg	0.69	0.68	0.69	525

```
Logistic Regression Training Score:
 81.03
Logistic Regression Test Score:
 68.38
```

Model-2(KNN):

Description:

Using supervised machine learning, the K-Nearest Neighbors (KNN) technique is used to solve regression and classification issues.

Hyperparameters:

```
knn_params={
    "n_neighbors":range(1,20,2),
    "weights":["uniform","distance"],
    "algorithm":["auto","ball_tree","kd_tree","brute"],
    "metric":["euclidean","minkowski","manhattan"],
    "leaf_size":range(1,30,5)
}
from sklearn.model_selection import GridSearchCV,RepeatedStratifiedKFold
grids=GridSearchCV(estimator=model,param_grid=knn_params,n_jobs=1,cv=3,scoring="accuracy",error_score=0)
res=grids.fit(X_train,y_train)
par_model=model.set_params(**res.best_params_)
```

Accuracy, Precision and Recall:

	precision	recall	f1-score	support
0	0.41	0.47	0.44	147
1	0.78	0.74	0.76	378
accuracy			0.66	525
macro avg	0.60	0.60	0.60	525
weighted avg	0.68	0.66	0.67	525

Model-3(Random Forest):

Description:

A random forest is a meta estimator that employs averaging to increase prediction accuracy and manage over-fitting after fitting several decision tree classifiers on different subsamples of the dataset. The best split strategy, or passing splitter="best" to the underlying DecisionTreeRegressor, is employed by the trees in the forest. If bootstrap=True (the default), the sub-sample size is managed using the max_samples argument; if not, each tree is constructed using the entire dataset.

Hyperparameters:

```
RandomForestClassifier(criterion='entropy', max_depth=15, max_features=0.75, min_samples_leaf=7, min_samples_split=3, n_estimators = 130)
```

Accuracy, Precision and Recall:

	precision	recall	f1-score	support
0	0.48	0.44	0.46	163
1	0.76	0.79	0.77	362
accuracy			0.68	525
macro avg	0.62	0.61	0.62	525
eighted avg	0.67	0.68	0.68	525

Initial Model Training Code, Model Validation and Evaluation Report

Initial Model Training Code:

```
# Importing modules
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,confusion_matrix
from sklearn import linear_model
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import Perceptron
from sklearn.tree import DecisionTreeClassifier
```

Output Screenshots

Hyperparameter tuning:

Model	Tuned Hyperparameters	Optimal Values
KNN	"n_neighbors", "weights", "algorithm", "metric", "leaf_size"	3,uniform,auto,euclidean,1
Random forest classifier	Criterion, max_depth, max_features, min_samples_leaf, min_samples_split, n_estimators	Entropy,15,0,75,7,3,130
Logistic regression	'C', 'max-iter', 'solver'	0.0001, 1000,lbfgs

Advantages:

1. Accuracy and Efficiency:
 - Machine learning classifiers can analyze large volumes of patient data efficiently, potentially identifying patterns and relationships that are not immediately apparent through traditional statistical methods.
 - This accuracy can lead to earlier detection of liver disease or risk factors, allowing for timely intervention and improved patient outcomes.
2. Personalized Medicine:
 - By leveraging patient records which include demographic information, medical history, and biomarkers, machine learning models can tailor predictions and treatments to individual patients.
 - This personalized approach enhances healthcare delivery by optimizing treatment plans and resource allocation based on patient-specific risks.
3. Integration with Electronic Health Records (EHRs):
 - Many healthcare systems now utilize electronic health records (EHRs) that contain a wealth of patient information.
 - Machine learning classifiers can seamlessly integrate with EHR systems, allowing for real-time analysis and decision support in clinical settings.
4. Feature Selection and Interpretability:
 - Machine learning models can automatically select relevant features from patient records, identifying key predictors of liver disease.
 - Techniques such as feature importance ranking in Random Forests or coefficient interpretation in Logistic Regression provide insights into which patient variables are most influential.
5. Scalability and Automation:
 - Once trained, machine learning models can automate the prediction process, reducing the burden on healthcare professionals and enabling scalable deployment across different healthcare facilities.

Disadvantages:

1. Data Quality and Preprocessing:
 - Patient records can be incomplete, inconsistent, or contain errors, which can affect the performance of machine learning models.
 - Extensive data preprocessing steps, such as handling missing values and standardizing data formats, are often required to ensure data quality.
2. Interpretability and Trust:
 - Complex machine learning models (e.g., deep learning) may provide high accuracy but lack interpretability, making it challenging to understand how predictions are made.
 - Clinicians may be reluctant to trust predictions from black-box models without clear explanations or validation against clinical guidelines.
3. Overfitting and Generalization:
 - Machine learning models, especially when trained on large and diverse patient datasets, may overfit to noise or specific characteristics of the training data.
 - Ensuring models generalize well to new patient data from different demographics or geographic regions is crucial for their clinical applicability.
4. Ethical and Legal Considerations:
 - Predictive models based on patient records raise ethical concerns related to patient privacy, informed consent, and potential biases in the data.
 - Adhering to regulations such as GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act) is essential to protect patient confidentiality and rights.
5. Clinical Integration and Validation:
 - Successfully integrating machine learning predictions into clinical workflows requires collaboration between data scientists, clinicians, and healthcare administrators.
 - Validating model predictions against real-world outcomes and ensuring they align with clinical guidelines are critical steps to gaining acceptance and adoption in healthcare settings.

3. MICROSOFT CERTIFIED:AZURE ADMINISTRATOR CERTIFICATE

Azure administrator implements, manages, and monitors various components including virtual networks, storage, compute, identity, security, and governance within the Azure environment:

Virtual Networks:

1. Implementation: Azure administrators are responsible for creating and configuring virtual networks (VNets) to provide secure and isolated network environments for Azure resources. This involves defining IP address ranges, subnets, and configuring network security groups (NSGs) to control traffic flow.
2. Management: Administrators manage virtual network settings, such as DNS settings, route tables, and network peering. They also oversee network connectivity, including virtual private network (VPN) gateways, Azure ExpressRoute, and network security solutions.
3. Monitoring: Azure administrators monitor virtual network performance, traffic patterns, and network security using Azure Monitor and Azure Network Watcher. They monitor network latency, traffic flows, and diagnose network issues to ensure optimal network performance.

Storage:

1. Implementation: Administrators implement Azure storage solutions, including Blob storage, Azure Files, Azure Queue storage, and Azure Table storage. They configure storage accounts, access keys, and encryption settings to store and manage data in the Azure cloud.
2. Management: Azure administrators manage storage accounts, implement storage replication for redundancy, and configure access control and permissions for data stored in Azure. They also optimize storage performance and cost by utilizing storage tiers and data lifecycle management.
3. Monitoring: Azure administrators use Azure Monitor and Azure Storage Analytics to monitor storage account performance, track data access patterns, and analyze storage usage for capacity planning and optimization.

Compute:

1. Implementation: Azure administrators provision and manage virtual machines (VMs) in Azure, including configuring VM sizes, storage, and networking settings. They also deploy and manage Azure App Services for hosting web applications and APIs.
2. Management: Administrators manage VMs, scale sets, and app services, including implementing availability sets, load balancing, and auto-scaling for high availability and performance. They also oversee container-based applications using Azure Kubernetes Service (AKS) and Azure Container Instances.
3. Monitoring: Azure administrators utilize Azure Monitor to monitor VM performance, resource utilization, and application performance. They also set up monitoring for containerized applications and services running in Azure Kubernetes Service using Azure Monitor and Application Insights.
4.

Identity:

1. Implementation: Azure administrators implement Azure Active Directory (AAD) to manage user identities, groups, and application access. They configure user authentication methods, multi-factor authentication, and self-service password reset.
2. Management: Administrators manage AAD users, groups, and roles to control access to Azure resources and applications. They also integrate on-premises Active Directory with Azure AD using Azure AD Connect for seamless identity management.
3. Monitoring: Azure administrators use Azure AD reporting and monitoring tools to track user sign-ins, security events, and access patterns. They also implement identity protection policies to detect and respond to identity-related security threats.

Security:

1. Implementation: Azure administrators implement security policies, role-based access control (RBAC), and Azure Security Center to secure Azure resources. They configure network security groups, security rules, and encryption for data at rest and in transit.
2. Management: Administrators manage security baselines, compliance policies, and threat detection settings using Azure Security Center. They also conduct security assessments, implement security best practices, and oversee security incident response.
3. Monitoring: Azure administrators monitor security alerts, vulnerabilities, and compliance status using Azure Security Center and Azure Monitor. They analyze security logs, investigate potential security threats, and implement security controls to mitigate risks.

Governance:

1. Implementation: Azure administrators implement governance policies using Azure Policy to enforce compliance and best practices across Azure resources. They define and enforce resource tagging, naming conventions, and access controls using Azure Policy.
2. Management: Administrators manage governance policies, compliance assessments, and remediation tasks using Azure Policy. They also implement resource locks, cost management controls, and organization using Azure Resource Manager.
3. Monitoring: Azure administrators monitor compliance status, policy violations, and resource configurations using Azure Policy and Azure. They track governance-related events, audit logs, and compliance trends to ensure adherence to organizational policies.

Overall, Azure administrators play a crucial role in, managing, and monitoring an organization's Azure environment, ensuring the efficient operation, security, and compliance of Azure resources and services. They leverage a range of Azure management and monitoring tools to maintain a robust and well-managed Azure environment that meets the organization's business and technical requirements.

The various features which can be configured are given below:

1. Azure Policy:
    - Azure administrators can configure Azure Policy to enforce governance and compliance requirements across the Azure environment. They define policy definitions and initiatives, assign policies to specific scopes (such as subscriptions, resource groups, or management groups), and enforce policy effects (e.g., auditing, denying non-compliant resources). This ensures that resources in Azure adhere to organizational standards, regulatory compliance, and security requirements.

2. Access Control:
    - Azure administrators manage access control using Azure Role-Based Access Control (RBAC). They assign Azure AD roles to users, groups, and applications, controlling access to Azure resources. Administrators define custom roles, grant permissions, and manage access at various levels within the Azure hierarchy, ensuring that users have appropriate access rights based on their roles and responsibilities.
3. Load Balancer:
    - Azure administrators configure Azure Load Balancer to distribute incoming network traffic across multiple VMs for high availability and scalability. They define load balancing rules, configure health probes, and set up backend pools to ensure even distribution of traffic. Additionally, they monitor load balancer performance, health probes, and traffic distribution using Azure Monitor to ensure optimal load balancing.
4. Storage Account and Blob Storage:
    - Administrators configure Azure Storage Accounts to store data in Azure. They create storage accounts, manage access keys, and configure encryption settings. Within storage accounts, administrators also configure Blob storage, defining containers, access permissions, and lifecycle management policies to optimize storage performance and cost-effectiveness. They monitor storage account usage, performance, and access patterns using Azure Storage Analytics and Azure Monitor.
5. Virtual Machine:
    - Azure administrators provision and manage virtual machines (VMs) in Azure, configuring VM sizes, storage, networking settings, and security. They implement availability sets, load balancing, and auto-scaling for high availability and performance. Additionally, they monitor VM performance, resource utilization, and application performance using Azure Monitor and set up monitoring for containerized applications and services running in Azure Kubernetes Service using Azure Monitor and Application Insights.
6. Monitor and Log Analytics:
    - Azure administrators use Azure Monitor to collect and analyze telemetry data from Azure resources. They configure monitoring solutions, set up alerts, and create dashboards to gain insights into resource performance, health, and usage. Additionally, administrators configure Log Analytics workspaces to centralize and analyze log and performance data from across the Azure environment, enabling advanced querying, reporting, and visualization of log data.

Configuring these components effectively enables Azure administrators to maintain a well-governed, secure, and efficient Azure environment, ensuring that resources adhere to best practices, compliance requirements, and operational standards. By leveraging Azure Policy, access control, network configuration, load balancing, storage management, virtual machine provisioning, and monitoring and log analysis tools, administrators can effectively manage and optimize the Azure environment to meet organizational needs.

4. CONCLUSION

The conclusion and future work for predicting liver diseases using logistic regression and decision tree classifier can be summarized as follows: Using logistic regression and decision tree classifier models, it is possible to predict the likelihood of liver diseases based on various risk factors and patient data. These machine learning algorithms can help identify high-risk individuals and assist healthcare professionals in making informed decisions for early diagnosis and intervention. From the study it is observed that as the training set ratio increases, the model's performance on the training data generally improves, as expected. This is indicated by the increasing trend in training accuracy. However, the performance on the testing data may not necessarily follow the same pattern. It may peak at a certain point and then start to decrease due to overfitting. A higher value of ccp_alpha increases the regularization strength, leading to simpler trees.

In this case, it is set to 0.04, indicating a moderate level of regularization. Managing the tree's depth aids in avoiding overfitting. While a deeper tree may be able to identify more complex patterns in the training set, overfitting could result from the tree learning to remember noise. The tree more effectively generalizes to unknown data by restricting the depth. Feature Selection and Model Optimization: Future work should focus on selecting the most relevant features that contribute significantly to the prediction of liver diseases. This can be achieved through feature selection techniques and model optimization. By reducing the number of input features, the models can be made more efficient and accurate. The future work includes: Ensemble Methods and Hybrid Models: Incorporating ensemble methods and hybrid models can further improve the predictive power of the logistic regression and decision tree classifier models.

By combining multiple models or algorithms, it is possible to achieve better accuracy and robustness in disease prediction. Incorporating Advanced Techniques: Exploring advanced machine learning techniques such as deep learning, random forests, and gradient boosting can lead to better disease prediction models. These techniques can capture complex patterns and relationships in the data, which may not be evident in logistic regression and decision tree classifier models. Handling Imbalanced Datasets: As liver diseases are relatively rare compared to other health conditions, datasets may be imbalanced.

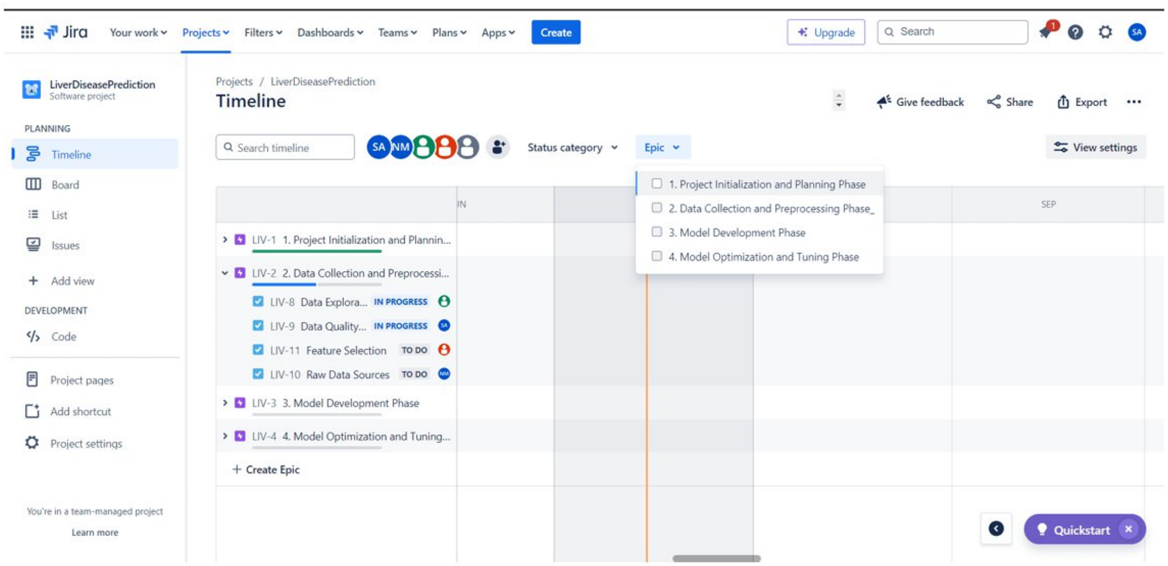
Handling Imbalanced Datasets: Future research should address techniques to handle imbalanced datasets,such as oversampling, undersampling, and synthetic minority oversampling technique (SMOTE), to improve the models' performance in predicting rare disease cases. Multi-class Prediction and Interpretability: Extending the current binary classification models to multi-class prediction can help identify various liver disease types. Additionally, ensuring the models' interpretability will allow healthcare professionals to understand the factors contributing to the disease prediction, leading to better decision-making and patient care. In conclusion, the prediction of liver diseases using logistic regression and decision tree classifier models has shown promising results. Further research and development in feature selection, model optimization, advanced techniques, real-world implementation, handling imbalanced datasets, and multi-class prediction will contribute to more accurate and reliable disease prediction models in the future.

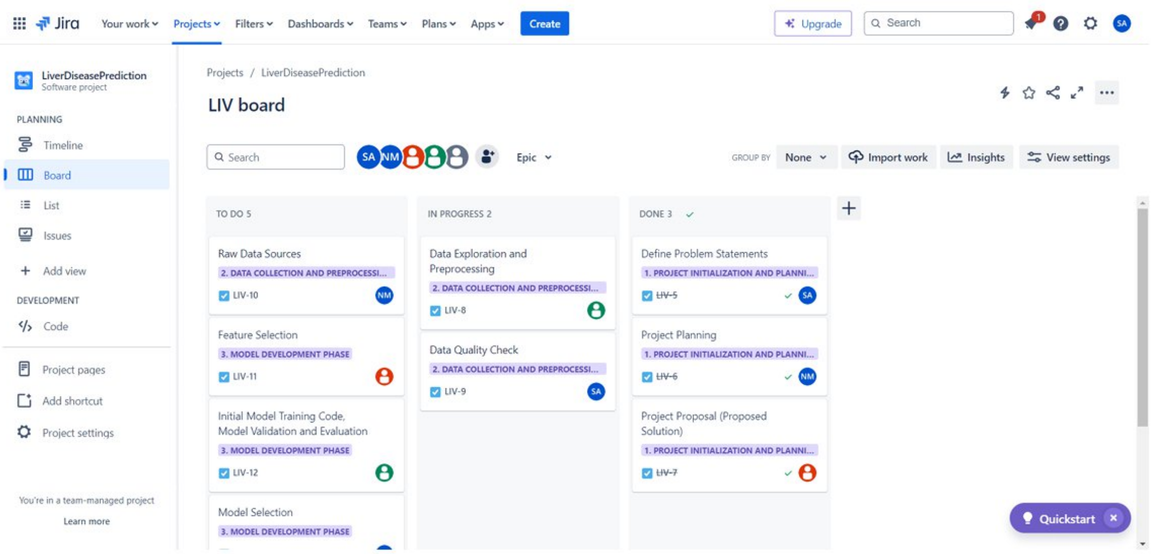
5.REFERENCES

1. Machine Learning Crash Course with TensorFlow APIs (Google)
https://developers.google.com/machine-learning/crash-course
2. “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/
3. TensorFlow in Practice Specialization (Coursera)
https://www.coursera.org/specializations/tensorflow-in-practice
4. Google Cloud AI and Machine Learning Products
https://cloud.google.com/products/ai
5. TensorFlow Official Website
https://www.tensorflow.org/
6. Microsoft Certified: Azure Administrator Associate
https://learn.microsoft.com/en-us/credentials/certifications/azure-administrator/?practice-assessment-type=certification

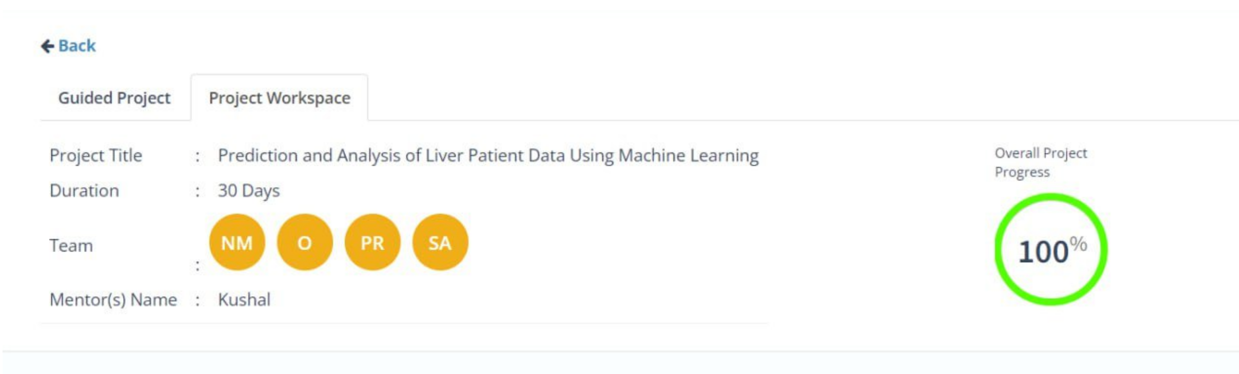
APPENDIX I

Sprint	Functional Requirement (Epic)	User Story Number	User Story	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	Data Collection and Preprocessing	LIV-5	Data Exploration and Understanding	1	High	Satwik	2/07/2024	5/07/2024
Sprint-1	Data Collection and Preprocessing	LIV-6	Data Pre-processing	2	High	Vigneshwar	2/07/2024	5/07/2024
Sprint-1	Data Collection and Preprocessing	LIV-7	Data Quality Check	1	Low	Naveen	2/07/2024	5/07/2024
Sprint-2	Model Development	LIV-8	Initial Model Training	1	Medium	Omprakash	6/07/2024	9/07/2024
Sprint-2	Model Development	LIV-9	Model Validation and Evaluation	1	High	Naveen	6/07/2024	9/07/2024





1. **Articial Intelligence using Google TensorFlow Marksheet**



2. Microsoft Azure AI 900 Examination Marksheet

Microsoft

EXAM SCORE REPORT

CANDIDATE	EXAM
Omprakash Pugazhendhi omprakash.2021@vitstudent.ac.in	**AI-900: Microsoft Azure AI Fundamentals** Registration ID: ITS-8506279 Exam reference #: 47404405 Date: July 25, 2024 ID: omprakash.2021@vitstudent.ac.in

RESULTS	100	200	300	400	500	600	700	800	900	1000
Required Score										
Your Score										

SECTION ANALYSIS	
Describe Artificial Intelligence Workloads and Considerations (15-20%)	87%
Describe Fundamental Principles of Machine Learning on Azure (20-25%)	76%
Describe Features of Computer Vision Workloads on Azure (15-20%)	91%
Describe Features of Natural Language Processing (NLP) Workloads on Azure (15-20%)	82%
Describe features of generative AI workloads on Azure (15-20%)	83%

FINAL SCORE	
Required Score	700
Your Score	822

OUTCOME	
Pass	✓

CERTIPORT