GUARD HER HEALTH

A WEB BASED BREAST CANCER PREDICTIOR

BY-
AYUSH AGARWAL:21104004
ABHILASH TYAGI:21104019
OJASVI PANDEY:21104027

PROJECT OVERVIEW:

"GUARD HER HEALTH" IS AN INNOVATIVE PROJECT DEDICATED TO TRANSFORMING BREAST CANCER PREDICTION USING CUTTING-EDGE TECHNOLOGY AND SPECIALIZED MACHINE LEARNING MODELS.

MISSION STATEMENT:

OUR MISSION IS TO PROVIDE ACCURATE AND ACCESSIBLE PREDICTIVE MODELS TAILORED FOR BREAST CANCER OCCURRENCE, REVOLUTIONIZING THE LANDSCAPE OF EARLY DETECTION AND PROACTIVE HEALTHCARE MANAGEMENT.

LITERATURE SURVEY

Name	Approach	Merits	Demerits
Significance of machine learning in healthcare: features, pillars and applications.	Examines the significance of machine learning in healthcare through the lens of features, pillars, and applications.	• Comprehensive exploration of machine learning's role in healthcare decision-making. • Highlights key features and pillars contributing to the success of machine learning in healthcare. • Discusses various applications, including disease diagnosis and personalized treatment.	• Limited emphasis on specific case studies or real-world implementations. • Does not deeply address potential challenges or limitations in the application of machine learning in healthcare.

Involvement of machine learning tools in healthcare decision making	Emphasizes the pivotal role of machine learning in swift and accurate healthcare decision-making across various domains.	• Highlights machine learning's significance in disease prediction, medical imaging, biomedicine, and drug discovery. • Acknowledges continuous enhancements in algorithms, particularly in deep learning for feature extraction.	• Lacks in-depth discussion on specific challenges faced in implementing machine learning tools in healthcare decision-making • Limited exploration of potential ethical considerations associated with machine learning in healthcare.

Breast cancer classification using machine learning: a comparative study	Conducts a thorough comparative study on the application of machine learning techniques for breast cancer classification.	• Comprehensive evaluation of multiple machine learning algorithms, providing a detailed overview of strengths and limitations. • Systematic analysis of classification accuracy, sensitivity, specificity, and roc metrics.	• Doesn't delve deeply into the nuances of algorithmic limitations. • Limited discussion on the real-world applicability of the findings in clinical settings.

Challenges in Breast Cancer Diagnosis

- **FALSE POSITIVES:**
  - CHALLENGE: HIGH OCCURRENCES OF FALSE POSITIVES LEAD TO UNNECESSARY MEDICAL PROCEDURES.
  - **CONSEQUENCE:** PATIENTS EXPERIENCE EMOTIONAL DISTRESS, UNNECESSARY TREATMENTS, AND A BURDEN ON HEALTHCARE RESOURCES.
- **DIAGNOSTIC PRECISION:**
  - **CHALLENGE:** EXISTING DIAGNOSTICS LACK THE PRECISION TO DISTINGUISH BETWEEN BENIGN AND MALIGNANT CASES ACCURATELY.
  - **CONSEQUENCE:** INACCURACIES MAY DELAY CRUCIAL TREATMENTS OR LEAD TO UNNECESSARY INTERVENTIONS.
- **RESOURCE STRAIN:**
  - **CHALLENGE:** OVER-RELIANCE ON LESS ACCURATE DIAGNOSTICS STRAINS HEALTHCARE RESOURCES.
  - **CONSEQUENCE:** SUBOPTIMAL RESOURCE ALLOCATION, INCREASED HEALTHCARE COSTS, AND PATIENT ANXIETY.

PROJECT FOCUS:

FOCUSED ON THE DEVELOPMENT OF SOPHISTICATED PREDICTIVE MODELS FROM SCRATCH AND THE CREATION OF AN INTUITIVE WEB PLATFORM, "GUARD HER HEALTH" AIMS TO EMPOWER BOTH HEALTHCARE PROFESSIONALS AND INDIVIDUALS IN MAKING INFORMED DECISIONS REGARDING BREAST CANCER.

KEY OBJECTIVES:

- ACCURACY IN BREAST CANCER PREDICTION
- ACCESSIBILITY AND USER-FRIENDLINESS
- REAL-LIFE IMPACT THROUGH EARLY DETECTION

PROBLEM STATEMENT AND MOTIVATION

PROBLEM STATEMENT:

- THE PERSISTENT ISSUE OF FALSE POSITIVES IN BREAST CANCER DIAGNOSIS LEADS TO UNNECESSARY MEDICAL INTERVENTIONS, CAUSING EMOTIONAL DISTRESS AND STRAINING HEALTHCARE RESOURCES

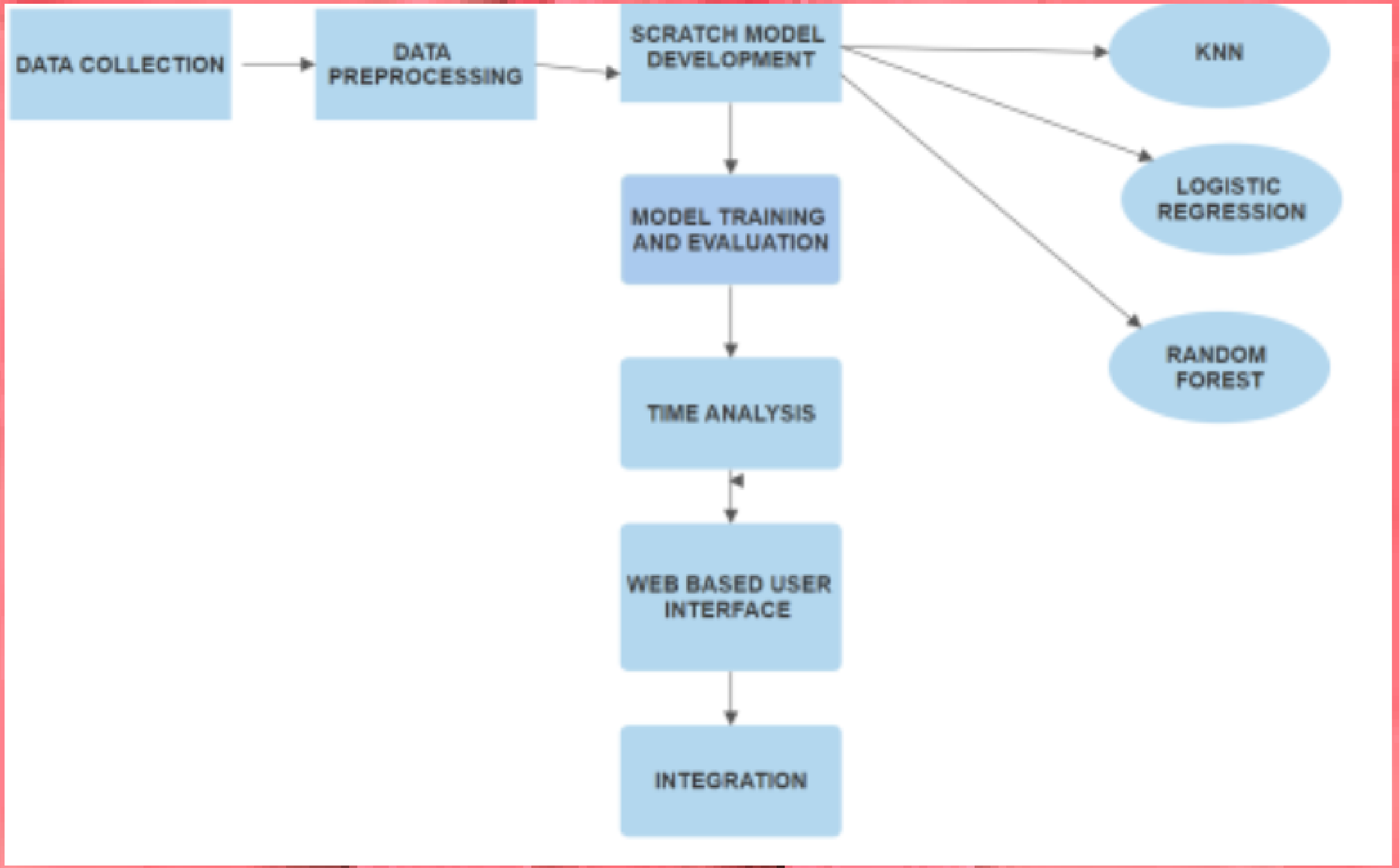
CORE MOTIVATION:

ADDRESSING THE FLAWS IN EXISTING DIAGNOSTICS AND MINIMIZING FALSE POSITIVES DROVE OUR ENDEAVOR TO CREATE A PRECISE AND DEPENDABLE PREDICTIVE TOOL FOR BREAST CANCER

.

ENHANCING ACCURACY IN DIAGNOSIS, REDUCING UNNECESSARY SURGERIES, AND EMPOWERING HEALTHCARE PROFESSIONALS WITH A RELIABLE TOOL FOR TIMELY INTERVENTIONS

WORKFLOW DIAGRAM



Overview of Data Collection and Cleaning Process

DATA COLLECTION:

- **DIVERSE SOURCES:** GATHERING EXTENSIVE MEDICAL DATA FROM REPUTABLE REPOSITORIES.
- **COMPREHENSIVE ARRAY:** INCLUSION OF VARIOUS FACTORS INFLUENCING BREAST CANCER OCCURRENCE.

DATA PREPROCESSING:

- **NORMALIZATION AND SCALING:** ENSURING UNIFORMITY AND USABILITY FOR MACHINE LEARNING MODELS.

Model Selection Rationale - Logistic Regression

- Dataset Suitability: The Wisconsin dataset involves predicting whether breast cancer tumors are malignant or benign, making it a binary classification task.

- Interpretability: Logistic Regression provides a clear understanding of the impact of individual features on the likelihood of tumor classification. For instance, it can reveal which features contribute most to distinguishing between malignant and benign tumors

Model Selection Rationale - KNN

Similarity-based Classification: Ideal for a dataset with varying tumor characteristics.

Adaptability: Doesn't assume a specific data distribution, suitable for diverse datasets.

Proximity in Feature Space: Effectively identifies similarities between instances, aiding classification

Model Selection Rationale - Random Forest

- Complexity Handling: Handles intricate relationships between features, ideal for non-linear patterns.
- Robustness: Maintains stability against overfitting, crucial for handling complex medical datasets.
- Enhanced Accuracy: Capable of capturing varied relationships within the dataset

PROJECT COMPONENTS

"GUARD HER HEALTH" WEBSITE OVERVIEW:

- **INTUITIVE INTERFACE:** SIMPLIFYING BREAST CANCER PREDICTION FOR USERS.
- **USER ENGAGEMENT:** EASY INTERACTION WITH PREDICTIVE MODELS.

FRONTEND AND BACKEND TECHNOLOGIES:

- **HTML, CSS, JAVASCRIPT:** BUILDING INTERACTIVE USER INTERFACES.
- **FLASK (BACKEND):** SEAMLESS INTEGRATION FOR USER REQUESTS AND DATA PROCESSING.

Novelty of the Project Approach

INNOVATIVE INTEGRATION:

- **WEB DEVELOPMENT & ML:** INTEGRATES WEB DEVELOPMENT AND ML FOR BREAST CANCER PREDICTION.
- **CUSTOM-BUILT ALGORITHMS:** DEVIATES FROM GENERIC MODELS WITH TAILOR-MADE ALGORITHMS.

POTENTIAL IMPACT:

- **TRANSFORMATIVE POTENTIAL:** RESHAPING THE LANDSCAPE OF BREAST CANCER DIAGNOSTICS.
- **EMPOWERING HEALTHCARE:** PROVIDING TOOLS FOR ACCURATE, USER-FRIENDLY DIAGNOSTICS.

Solution Approach

- **MODEL DEVELOPMENT OVERVIEW:**
  - **TAILORED PREDICTIVE MODELS:** DEVELOPMENT OF CUSTOM ML MODELS(RANDOM FOREST,KNN,LOGISTIC REGRESSION) FOR BREAST CANCER PREDICTION.
- **TECHNOLOGY-HEALTHCARE INTEGRATION:**
  - **FUSION OF DISCIPLINES:** BLENDING TECH INNOVATION WITH HEALTHCARE PRECISION.
  - **USER-CENTRIC DESIGN:** INCORPORATING MODELS INTO AN INTUITIVE, ACCESSIBLE WEB INTERFACE.

Technologies Leveraged

- **HTML, CSS, JAVASCRIPT:**
  - **HTML:** FOUNDATION FOR STRUCTURING USER INPUT FORMS.
  - **CSS:** ENHANCES USER EXPERIENCE, MAKING THE INTERFACE VISUALLY APPEALING.
  - **JAVASCRIPT:** FACILITATES DYNAMIC INTERACTIONS, FORM VALIDATION, AND USER ENGAGEMENT.
- **FLASK INTEGRATION:**
  - **BACKEND FUNCTIONALITY:** ORCHESTRATES DATA PROCESSING, INTEGRATES MACHINE LEARNING MODELS.
  - **ROUTING AND VIEWS:** FACILITATES SEAMLESS USER INTERACTION AND MODEL INVOCATION

Integrating Models with the Website

- **MODEL DEPLOYMENT:**
  - **SEAMLESS INTEGRATION:** IMPLEMENTING MODELS WITHIN THE "GUARD HER HEALTH" WEBSITE.
  - **USER INTERACTION:** ENABLING USERS TO INTERACT WITH PREDICTIVE MODELS THROUGH THE WEBSITE.
- **FUNCTIONALITY HIGHLIGHTS:**
  - **USER-FRIENDLY INTERFACE:** MAKING PREDICTIONS EASILY ACCESSIBLE TO USERS.
  - **REAL-TIME INSIGHTS:** PROVIDING INSTANT BREAST CANCER LIKELIHOOD ESTIMATIONS.

Conclusion

Significant Achievements and Impact

Summary of Achievements:

- Developed Specialized Models: Logistic Regression, Random Forest, KNN.
- Integrated Technologies: HTML, CSS, JavaScript, Flask for website creation.
- Impactful Predictions: Providing insights into breast cancer likelihood

Impact on Breast Cancer Diagnosis:

- Enhanced Prediction Accuracy: Offering informed insights for early detection.
- Empowered Decision-Making: Enabling proactive healthcare management

Impact on Healthcare

- Empowering Proactive Healthcare: Early detection for better outcomes.
- User-Centric Healthcare: Assisting healthcare professionals and individuals.

Future Work

Potential Areas for Further Improvement

- **Advanced Feature Engineering:**
  - **Explore sophisticated techniques for enhanced interpretability.**

- Future Work: Integration of Chatbot for Enhanced User Interaction

Enhancement Opportunities for the Project

- **Ensemble Learning Strategies:**
  Experiment with ensemble methodologies for improved accuracy.

- **Data Expansion:**
  Augment the dataset to ensure model robustness.


THANK
YOU