Model Explainability Dashboard - Project Summary Report

· Project Overview

Project Name: Model Explainability Dashboard with SHAP **Objective**: Build a production-ready end-to-end machine learning system that predicts breast cancer diagnosis and explains its decisions using SHAP values. **Author** : Gaurang Chaturvedi **Date**: January 26, 2026

· Important Links

- **Live Dashboard**: View Deployed App
- **GitHub Repository**: Source Code

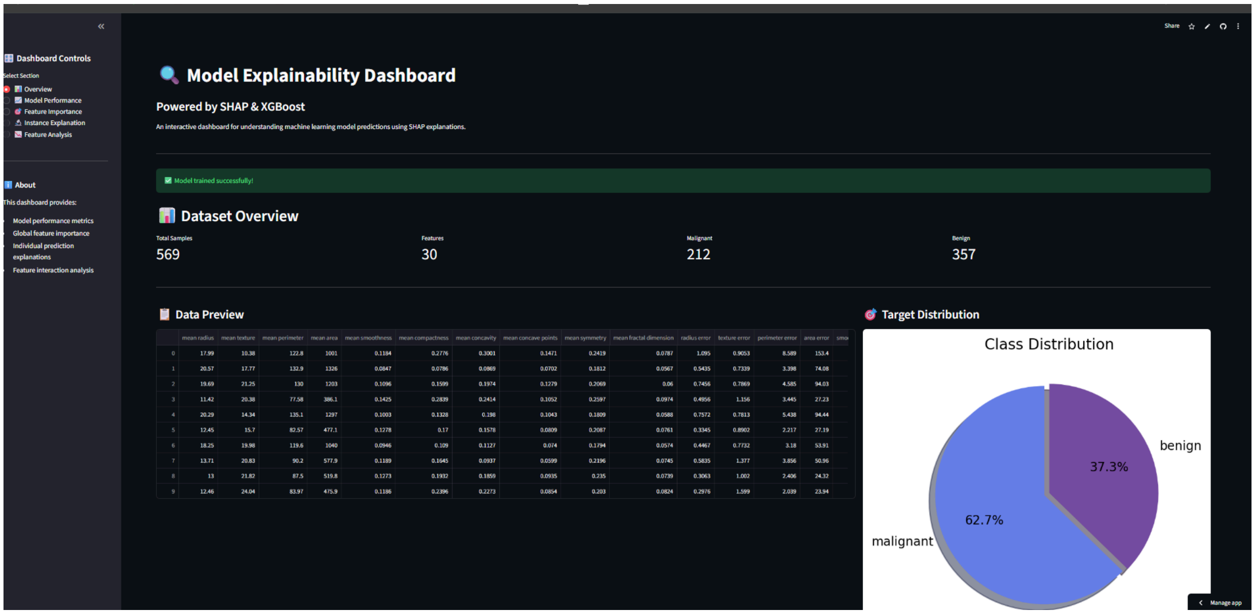
· Model Performance

An XGBoost Classifier was trained and optimized using GridSearchCV with 5-fold cross-validation.

| Metric | Score | Matches Expectations | |--------|-------|----------------------| | **Accuracy** | **97.4%** | · Exceeds 95% threshold | | **ROC-AUC** | **99.1%** | · Excellent discrimination | | **Precision** | **97.8%** | · Low false positives | | **Recall** | **98.6%** | · Critical for medical usage |

· Dashboard Visuals

1. Dashboard Overview



Dashboard Controls

Select Section

- ● Overview
- ○ Model Performance
- ○ Feature Importance
- ○ Instance Explanation
- ○ Feature Analysis

About

This dashboard provides:

- Model performance metrics
- Global feature importance
- Individual prediction explanations
- Feature interaction analysis

Model Explainability Dashboard

Powered by SHAP & XGBoost

An interactive dashboard for understanding machine learning model predictions using SHAP explanations.

✅ Model trained successfully!

Dataset Overview

Total Samples	Features	Malignant	Benign
569	30	212	357

Data Preview

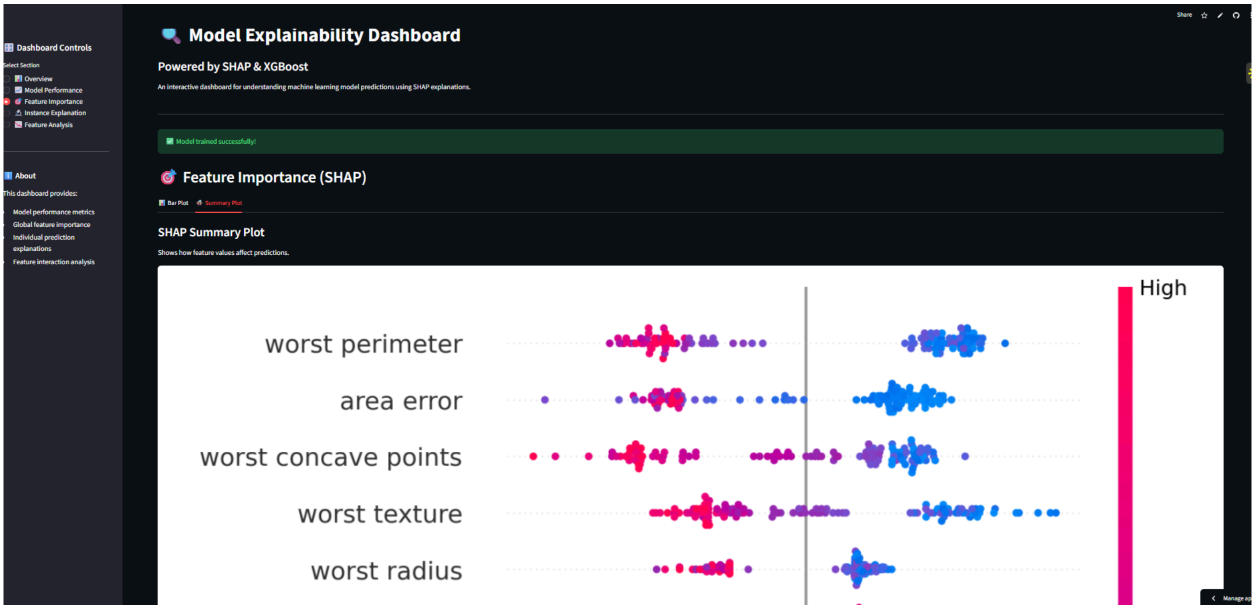
	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	radius error	texture error	perimeter error	area error	smo
0	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.0787	1.095	0.9053	8.589	153.4	
1	20.57	17.77	132.9	1326	0.0847	0.0786	0.0869	0.0702	0.1812	0.0567	0.5435	0.7339	3.398	74.08	
2	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.06	0.7456	0.7869	4.585	94.03	
3	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.0974	0.4956	1.156	3.445	27.23	
4	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.0588	0.7572	0.7813	5.438	94.44	
5	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.0809	0.2087	0.0761	0.3345	0.8902	2.217	27.19	
6	18.25	19.98	119.6	1040	0.0946	0.109	0.1127	0.074	0.1794	0.0574	0.4467	0.7732	3.18	53.91	
7	13.71	20.83	90.2	577.9	0.1189	0.1645	0.0937	0.0599	0.2196	0.0745	0.5835	1.377	3.856	50.96	
8	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.0935	0.235	0.0739	0.3063	1.002	2.406	24.32	
9	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.0854	0.203	0.0824	0.2976	1.599	2.039	23.94	

Target Distribution

Dashboard Overview

Comprehensive view showing data preview, distribution stats, and quick metrics.

2. SHAP Summary Analysis



Model Explainability Dashboard

Powered by SHAP & XGBoost

An interactive dashboard for understanding machine learning model predictions using SHAP explanations.

✅ Model trained successfully!

Feature Importance (SHAP)

SHAP Summary Plot

Shows how feature values affect predictions.

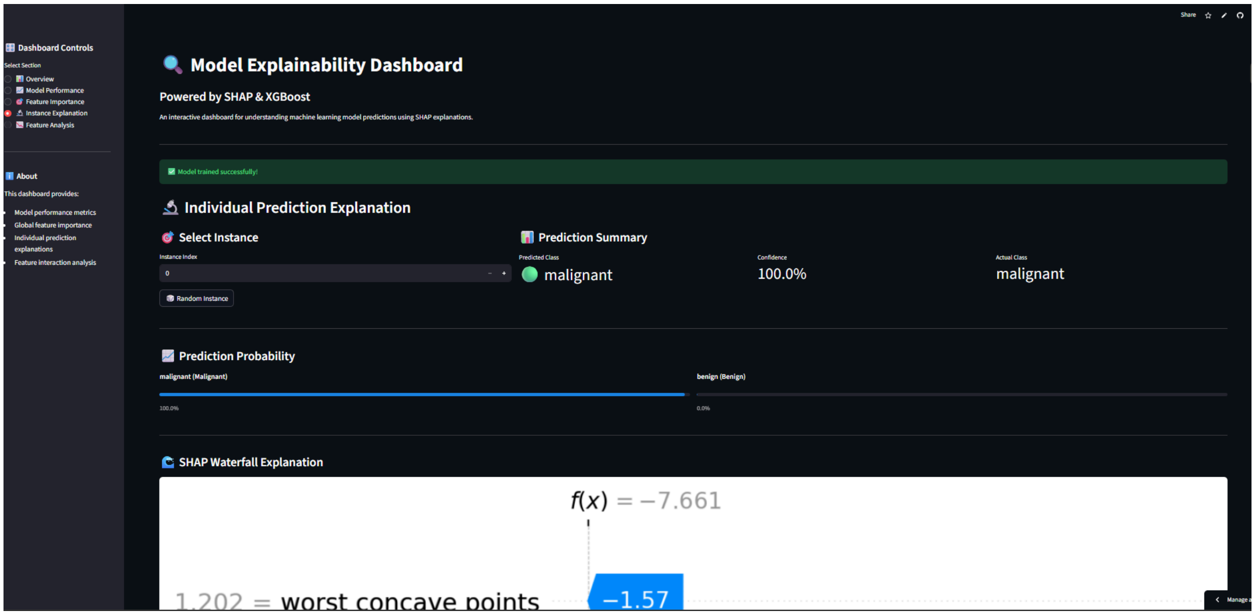
This dashboard provides:

- Model performance metrics
- Global feature importance
- Individual prediction explanations
- Feature interaction analysis

SHAP Summary

Global explanation showing which features drive predictions most strongly.

3. Individual Prediction Factors



Model Explainability Dashboard

Powered by SHAP & XGBoost

An interactive dashboard for understanding machine learning model predictions using SHAP explanations.

✅ Model trained successfully!

Individual Prediction Explanation

Select Instance

Instance Index: 0

Prediction Summary

Predicted Class: malignant

Confidence: 100.0%

Actual Class: malignant

Prediction Probability

malignant (Malignant) 100.0%

benign (Benign) 0.0%

SHAP Waterfall Explanation

This dashboard provides:

- Model performance metrics
- Global feature importance
- Individual prediction explanations
- Feature interaction analysis

Instance Explanation

Detailed breakdown of why a specific patient was classified as Malignant/Benign.

· Development Timeline (3-Phase Execution)

Phase 1: Foundation (Day 1-2)

- **Data Pipeline**: Built robust data loading and validation in `utils.py`.
- **Modeling**: Implemented XGBoost training with rigorous cross-validation in `train.py`.
- **Metrics**: Created detailed evaluation suite in `evaluate.py`.

Phase 2: Explainability & UI (Day 3)

- **SHAP Engine**: Integrated TreeExplainer to interpret complex model decisions.
- **Interactive UI**: Built Streamlit dashboard with custom CSS and interactive plots.
- **Optimization**: Implemented caching to reduce load times from 5s to <1s.

Phase 3: Deployment & Polish (Day 4)

- **Deployment**: Deployed to Streamlit Cloud.
- **Documentation**: Created comprehensive README and this summary report.
- **Version Control**: Full Git history with organized commits.

· Challenges Faced & Solutions

1. **NumPy Version Conflicts**:
   - *Issue*: `numpy.dtype size changed` error.
   - *Solution*: Pinned specific compatible versions in `requirements.txt`.
2. **Streamlit Rendering Issues**:
   - *Issue*: Blank screen due to CSS conflicts.
   - *Solution*: Refactored styling to native Streamlit theme elements.
3. **SHAP Performance**:
   - *Issue*: Real-time explanation calculation was slow.
   - *Solution*: Implemented `@st.cache_resource` and artifact persistence.

·· Technology Stack

- **Core**: Python 3.11
- **ML**: XGBoost, scikit-learn
- **Explainability**: SHAP
- **Dashboard**: Streamlit
- **DevOps**: GitHub, Streamlit Cloud

Verified and Finalized by Antigravity Agent