BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

WORK INTEGRATED LEARNING PROGRAMMES (WILP)

ABSTRACT / OUTLINE REPORT

Dissertation / Project / Project Work Title:

Automated Machine Learning (AutoML) for Healthcare Risk Prediction

Course No.: DSECLZG628T

Course Title: Dissertation / Project / Project Work

Dissertation / Project /Project Work Done by:

Student Name: Keerthi Kumar B
BITS ID: 2023DC04097
Degree Program: M.Tech. in Data Science & Engineering
Research Area: Health Care Management
Dissertation / Project Work carried out at: Individual Research Project (Executed Using Personal Computational Resources)

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

VIDYA VIHAR, PILANI, RAJASTHAN - 333031.

November 2025

Contents

1. Broad Area of Work

The project lies in the intersection of **Machine Learning, Artificial Intelligence, and Healthcare Analytics**. It focuses on leveraging **Automated Machine Learning (AutoML)** techniques to build predictive models for identifying patients at high risk of developing critical health conditions.

Key research areas include:

- **Data Preprocessing and Feature Engineering:** To prepare healthcare datasets for modeling.
- **AutoML Frameworks:** To automate model selection, hyperparameter optimization, and ensemble generation.
- **Explainable AI (XAI):** To interpret and visualize the decision factors influencing model predictions.
- **Model Deployment:** To create an end-to-end pipeline and interactive dashboard for prediction and interpretation.

2. Background

The healthcare industry generates massive volumes of patient data daily, but turning this data into actionable insights requires specialized data science expertise. Traditional model development is time-consuming and heavily dependent on domain experts and data scientists.

Automated Machine Learning (AutoML) provides a scalable, efficient, and accurate approach by automating the processes of model building, hyperparameter tuning, and performance evaluation.

This project addresses the growing need in healthcare for **data-driven decision support systems** that can quickly identify patients at risk of diseases such as stroke or heart disease. By integrating explainability tools, the system ensures **transparency and trust**, which are crucial in clinical settings.

3. Objectives

The main objectives of the project are as follows:

- To design and implement an automated machine learning pipeline for healthcare risk prediction.
- To compare the performance of traditional manually built models with AutoML-generated models.
- To apply explainability techniques (e.g., SHAP or LIME) to interpret model outcomes.
- To build a user-friendly interface (Streamlit Dashboard) for clinicians to visualize predictions and feature importance.
- To evaluate the models on performance metrics such as accuracy, ROC-AUC, and interpretability.

4. Scope of Work

The scope of this dissertation includes:

- Independent data preprocessing and exploratory data analysis using Python and Tableau.
- Manual model development (e.g., Logistic Regression, Random Forest) for baseline comparison.
- Implementation of AutoML frameworks such as **PyCaret**, **H2O AutoML**, or **Auto-sklearn**.
- Model explainability and fairness assessment using **SHAP** or **LIME**.
- Deployment of the final model as an interactive web application for demonstration purposes. All implementation, experimentation, and report documentation will be done independently by the student.

5. Plan of Work

Phases	Start Date-End Date	Work to be done
Dissertation Outline & Dataset Selection	01 Nov 2025 – 15 Nov 2025	Literature Review, Dataset Selection, and Outline Preparation
Data Preprocessing & EDA	16 Nov 2025 – 05 Dec 2025	Data cleaning, handling missing values, feature engineering, Tableau-based EDA
Manual Model Development	06 Dec 2025 – 20 Dec 2025	Implement baseline models and tune parameters
AutoML Implementation	21 Dec 2025 – 10 Jan 2026	Configure AutoML frameworks, evaluate best models, and compare results
Explainability & Fairness Analysis	11 Jan 2026 – 25 Jan 2026	Apply SHAP/LIME, evaluate feature influence and fairness
Deployment & Report Preparation	26 Jan 2026 – 15 Feb 2026	Streamlit dashboard, documentation and final report

6. Literature References

1. Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). *Efficient and robust automated machine learning*. Advances in Neural Information Processing Systems (NIPS).
2. Thornton, C., Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2013). *Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms*. KDD.
3. Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions*. Advances in Neural Information Processing Systems.
4. Rajkomar, A., Dean, J., & Kohane, I. (2019). *Machine learning in medicine*. New England Journal of Medicine, 380(14), 1347–1358.
5. Choudhury, A., & Asan, O. (2020). *Role of artificial intelligence in patient safety outcomes: Systematic literature review*. JMIR Medical Informatics, 8(7), e18599.

7. Particulars of the Supervisor and Examiner

	Supervisor	**Additional Examiner**
Name	Tejas Kambale	Shashank S P
Qualification	Bachelor of Engineering in Electronics and Communications	Bachelor of Engineering in Mechanical Engineering
Designation	Senior Software Engineer	Senior Engineer
Employing Organization and Location	Bosch Global Software Technologies, Bangalore	Bosch Global Software Technologies, Bangalore
Phone No.(with STD Code)	+91 9901400775	+91 9743523262
Email Address	tejaskambale07@gmail.com	shashankp2324@gmail.com

8. Remarks of the Supervisor

"The proposed dissertation on **"AutoML for Healthcare Risk Prediction"** is highly relevant to current advancements in artificial intelligence and healthcare analytics. The student has identified a meaningful application area that combines data-driven automation, transparency, and interpretability—key elements in ethical AI research. The project plan is well-structured and covers all aspects of model development, validation, and deployment. It holds strong potential for real-world impact and possible publication in the area of healthcare data science."

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
WORK INTEGRATED LEARNING PROGRAMMES (WILP) DIVISION
SECOND SEMESTER OF ACADEMIC YEAR 2025-2026
DSECLZG628T: Dissertation / Project / Project Work

STUDENT ID No.	2023DC04097
NAME OF THE STUDENT	Keerthi Kumar B
STUDENT'S EMAIL ADDRESS	2023dc04097@wilp.bits-pilani.ac.in
STUDENT'S EMPLOYING ORGANIZATION & LOCATION	Bosch Global Software Technologies, Bangalore
SUPERVISOR'S NAME	Tejas Kambale
SUPERVISOR'S EMPLOYING ORGANIZATION & LOCATION	Bosch Global Software Technologies, Bangalore
SUPERVISOR'S EMAIL ADDRESS	tejaskambale07@gmail.com
ADDITIONAL EXAMINAER'S NAME	Shashank S P
ADDITIONAL EXAMINER'S EMPLOYING ORGANIZATION & LOCATION	Bosch Global Software Technologies, Bangalore
ADDITIONAL EXAMINER'S EMAIL ADDRESS	shashankp2324@gmail.com
DISSERTATION / PROJECT / PROJECT WORK TITLE	Automated Machine Learning (AutoML) for Healthcare Risk Prediction

---

Please prepare the outline as a separate document with the following sections along with the above identification information.

1. Cover Page with Student ID No., Name, Course Number, Course Title and Dissertation / Project / Project Work Title, Broad Academic Area of Work.
2. Background (Relevance of the Project to the current work environment in the employing organization)
3. Objectives
4. Scope of Work (To be done by the student independently)
5. Plan of Work (Work to be done during the semester)
6. Literature References
7. Particulars of the Supervisor and Additional Examiner
8. Remarks of the Supervisor

Keerthi Kumar B	Tejas Kambale	Shashank S P
Signature of Student	**Signature of Supervisor**	**Signature of Additional Examiner**