



# AAM IPL 5

## Diagnosing Breast Cancer

B.Tech – CSE(AIML)

V Semester - ML and AIUP, Aug-Oct 2024

Department of Computer Science Engineering – AI and ML (CSM)

G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP

### Algorithm of Application

SVM – Support Vector Machines

### Project Title

Diagnosing breast cancer based on the diagnostic data

### Project Objective

This project implements a Support Vector Machine (SVM) classifier to predict the class of breast cancer (breast cancer detected or not detected) ) from the provided dataset.

The dataset is loaded from sklearn. The breast cancer dataset consists of 569 samples, each representing a patient with a set of features.

The dataset is used for training a SVM classifier and for evaluating performance metrics such as accuracy, precision, recall, F1-score.

### Dataset

- Description:
  - Implements a Support Vector Machine (SVM) – linear and RBF – classifiers to predict the class of breast tumor (breast cancer) from the provided dataset – malignant or benign.
  - The dataset needs to be loaded from scikit-learn
  - The dataset should be used for training a SVM classifier and evaluate its performance using various metrics such as accuracy, precision, recall, F1-score.
- Dataset Details:
  - The breast cancer dataset consists of 569 samples, each representing a patient with a set of features
- Data Source/Published By:
  - [Breast Cancer Wisconsin \(Diagnostic\) - UCI Machine Learning Repository](#)
  - [\[PDF\] Nuclear feature extraction for breast tumor diagnosis | Semantic Scholar](#)
  - Supported by CS Department, University of Wisconsin-Madison
- Features & Data Download Link



○ [Breast Cancer Wisconsin \(Diagnostic\) - UCI Machine Learning Repository](#)

## Implementation Steps

1. Define a method to add AAM-IPL water mark image (image provided along with shell code)
2. Load the breast cancer data from scikit-learn module
3. Print the dataset description/meta data
4. Print feature names, total data set/rows count, count of benign (non-cancerous)/malignant (cancerous) tumor in the data samples
5. Standardize, train and predict/infer the tumor class/type (malignant or benign)
6. Evaluate the model using various applicable metrics
7. Plot comparative plots of
  - a. Model accuracies
  - b. Confusion matrices
  - c. RoC curves
  - d. Precision-recall
  - e. F1 score
  - f. Decision boundaries
  - g. Feature Importance

## Project Files Provided

- Project shell code file - **AAM-IPL-Wk-5-SVM-ProjectName-Shell-Code-V3.ipynb**
- Training Data – **To be loaded from scikit-learn**
- Watermark image for plots - **AAM-IPL-Watermark-for-Plots.png**

## Project Overview, Implementation and Submission Timeline

19-10-2024 – Saturday 10:30 AM – Next Week Project Details Announcement – Topic, Data Set, Shell Code etc. Announcement Channels – Google Class, Industry Projects WhatsApp Group.			
8	19-10-2024 - Saturday Duration: 1.5 Hrs	Support Vector Machines (SVM) – Model Overview/Recap, Project Description, and Interactive Q&A	Online – Google Class
9	20-10-2024 - Sunday Duration: 1.5 Hrs	Support Vector Machines (SVM) – Model Building, Output Demonstration, Q&A	Online – Google Class
24-10-2024 – Thursday 11:59 PM - Deadline to upload the project code submission by all students in Google Class.			

### Guest Lecture Timings:

Saturdays: 10:30 AM IST – 12:00 Noon IST

Sundays: 10:30 AM IST – 12:00 Noon IST

Mondays: 6:30 PM IST – 8:00 PM IST

## Development Environment

- Computing Language – Python
- IDE – Visual Studio Code with Jupyter Notebook

## Instructor

Instructor

Venkateswar Reddy Melachervu (alumnus of  
GPREC, ECE Class of '92)  
CTO, Brillium Technologies, Bengaluru  
Email: [venkat.reddy.gf@gprec.ac.in](mailto:venkat.reddy.gf@gprec.ac.in)  
Profile: [LinkedIn](#)

Visiting Faculty

## Coordination

All the activities of this programme – lecture venues, weekly projects details announcements, general announcements, changes in lecture timings, etc. will be coordinated by CSM faculty member Sri V.Suresh.

Channels of Communication and Announcements

- Google Classroom
- Whatsapp group - **Applied AI & ML Industry Projects Lab**
- Emails (Strictly GPREC email addresses only)

## Programme Coordinator

Prof. V.Suresh  
Email: [vsuresh.ecs@gprec.ac.in](mailto:vsuresh.ecs@gprec.ac.in)

Faculty Member, CSM

## Reference Books

- Pattern Recognition and Machine Learning by Chris Bishop, 2006 – [PDF Link](#)
- Machine Learning using Python by Manaranjan Pradhan and U Dinesh Kumar, Wiley 2019 – [PDF Link](#)

## Policies

- **Attendance:** All sessions are expected to be attended by all the enrolled students. In case of inability to attend, prior information is expected to be provided by the student to the coordinator with a copy to the visiting faculty
- **Project Submissions:** Duly completed projects (Jupyter NB file and a PDF of the Jupyter NB file) are expected to be submitted through google class prior to the deadline. In case of inability to complete due various unforeseen circumstance, students are expected to seek extension for the submission deadline.
- **Academic Integrity:** Students are expected to uphold the highest standards of academic integrity in all assignments for the Applied AI & ML Industry Projects Lab. Each assignment must be the student's own work, and all sources and collaborators must be properly acknowledged. By submitting their completed project source code, students confirm that they have adhered to this integrity policy and completed their work in an honest and ethical manner.

## Additional Information

For students interested in engaging with special projects in the field of Gen AI, please reach out to the visiting faculty at [venkat@brillium.in](mailto:venkat@brillium.in) for further details and opportunities.

## Contact Information

For any questions or concerns or further details on this programme, please contact **Program Coordinator** during office hours or via email.

----- End of the Document -----