

S3 MCA (2024-26 Batch)

M24CA1L306 Data Science Lab

Lab Exercise #07

Date of Submission: 26 Sept 2025

- Download the dataset [titanic.csv](#)
- Choose the appropriate k-Value
- Do the following tasks:
 1. [Implement the k-NN Algorithm](#) to identify passenger survival based on demographics and travel information.
 2. Demonstrate how does the choice of k affect survival prediction accuracy?
 - Does a small k (e.g., 1–3) overfit to the training data?
 - Does a larger k smooth out decision boundaries too much?
 3. Does feature scaling improve the performance of KNN on Titanic data?
 - Compare accuracy with and without normalization/standardization.
 4. What does the confusion matrix reveal about false positives and false negatives in survival predictions?
 - Which error type is more common?
 5. How stable are KNN predictions across different train-test splits?

S3 MCA (2024-26 Batch)

M24CA1L306 Data Science Lab

Lab Exercise #08

Date of Submission: 26 Sept 2025

- Download the dataset [Breast_Cancer.csv](#)
- Choose the appropriate k-Value
- Do the following tasks:
 1. [Implement the k-NN Algorithm](#) to detect malignant vs benign tumors using cell nucleus features.
 2. How does the choice of k influence the classification accuracy?
 - Is there an optimal k that balances bias and variance?
 3. Which class (benign vs malignant) is classified more accurately by KNN?
 - Inspect the confusion matrix for per-class performance.
 4. Does feature scaling (standardization/normalization) significantly impact KNN results?
 - Compare accuracy before and after scaling.
 5. Which features (e.g., mean radius, texture, smoothness) have the most influence on classification accuracy?
 - Train KNN using subsets of features.
 6. What is the effect of train-test split ratio (e.g., 60-40 vs 80-20) on KNN performance stability?