K-Nearest Neighbors

—

By Zach Jordan and Logan Liddiard

Topics

Part 1: Heart Disease

Part 2: Forest Fires

Heart Disease Dataset

❖ Values
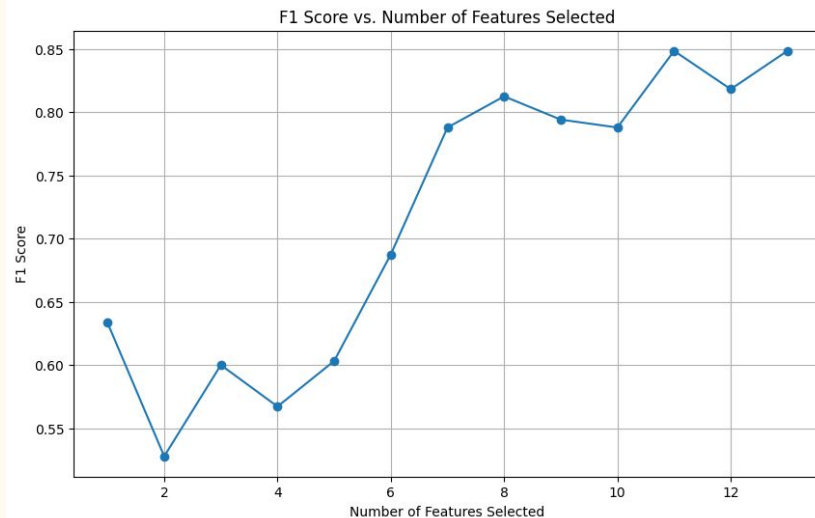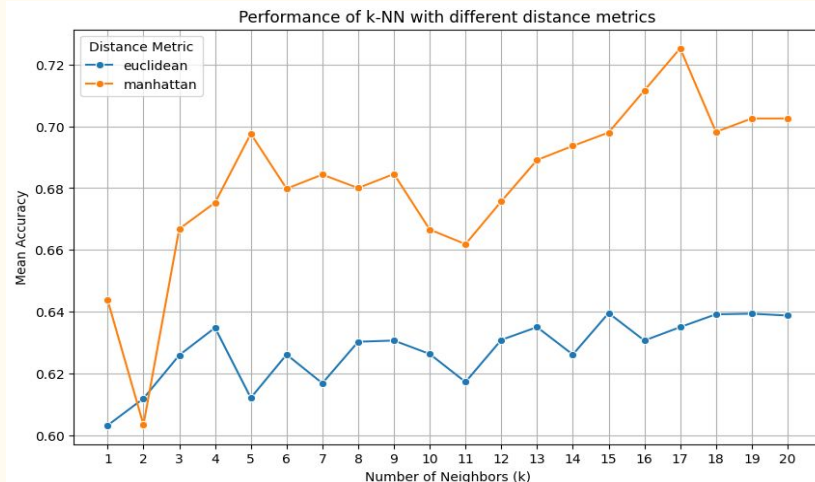 ➢ Patient information
  ➢ Health data
  ➢ Heart stats

❖ Cleanup
  ➢ Only used integer values
  ➢ Renamed values for readability
  ➢ Created a boolean disease value

Heart Disease

❖ Best combos for high F1 scores
  ➢ 11-13 values
❖ Mean Precision
  ➢ 0.693
❖ Mean Recall
  ➢ 0.615
❖ Mean F1 Score
  ➢ 0.637

Performance of k-NN with different distance metrics

F1 Score vs. Number of Features Selected

Forest Fires Dataset

❖ Values
  ➢ Location values
  ➢ Meteorology data
  ➢ Fire Index data
❖ Cleanup
  ➢ Checked if area was burned
  ➢ Took out months and days

ONLY YOU
CAN PREVENT WILDFIRES

Forest Fires

❖ Best combos for high F1 scores
  ➢ Relative Humidity
  ➢ Temperature
  ➢ Duff Moisture Code index from the Fire Weather Index system
  ➢ Drought Code index from the Fire Weather Index system.
❖ Mean Precision
  ➢ 0.650
❖ Mean Recall
  ➢ 0.671
❖ Mean F1 Score
  ➢ 0.651