KNN Project Proposal

Zara Tekmen & Margaret Jackson

Potter IB CS HL

**Project Description**

The "Dry Bean Dataset" describes the major characteristics of seven different types of beans. The measurements are taken from high resolution photos and all measurements are in pixels. The data has 16 features as well as the bean classification.

[Kaggle Link](https://www.kaggle.com/datasets/nimapourmoradi/dry-bean-dataset-classification?resource=download) [Google Sheet Data Set](https://docs.google.com/spreadsheets/d/1Qvdfe_vF_yP0ZS38rSaT9ceqeNR6aMiR2mXTAC8gQ0g/edit?usp=sharing)

**Data Values:**

* Area: Integer value measuring the amount of pixels contained within the photo of the bean
  + Integer values from 20k to 255k pixels
* Perimeter: bean circumference defined by the length of its border
  + Decimal Numbers from 525 to 1990
* Major Axis Length: distance of the longest straight line that can be drawn from a bean
  + Decimal Numbers from 184 to 739
* Minor Axis Length: distance of the longest line that is drawn perpendicular to the major axis
  + Decimal Numbers from 123 to 460
* Aspect Ration: relationship between the major axis and minor axis
  + Decimal Numbers from 1.02 to 2.43
* Eccentricity: the eccentricity of the ellipse that has the same moments as the bean. The moments are the end points of the major and minor axes
  + Decimal Numbers from 0.22 to 0.91
* Convex Area: number of pixels in the smallest convex polygon that contains the area of a bean seed
  + Integers from 20700 to 26300
* Equivalent Diameter: diameter of a circle with the same area as the bean
  + Decimal Numbers from 161 to 569
* Extent: the ratio of the pixels in the bounding box around the bean to the bean area
  + Decimal Numbers from 0.56 to 0.87
* Solidity: the ratio of the pixels in the convex shell of the bean to those that are contained within the bean
  + Double from 0.92 to 0.99
* Roundness: the formula is 4 𝛑 multiplied by Area divided by Perimeter squared
  + Decimal Numbers from 0.49 to 0.99
* Compactness: measures the roundness of an object
  + Decimal Numbers from 0.64 to 0.99
* 4 Shape Factors: The shape factor is not described and the range of values is small so we aren’t using this data.
* (**Target**) Class: String value representing the type of dried bean photographed and measured
  + String value with 7 types: Seker, Barbunya, Bombay, Cali, Horoz, Sira, and Dermason