Paul Robles 04/20/2023

**Lab 2: Classification Problem (RESUBMISSION)**

**Abstract:**

The purpose of this study is to be address the problem of predicting camera price categories of low, medium and high based on their features. The models that I used were logistic regression, K-Nearest Neighbors and decision tree. The decision tree had the highest accuracy of 67% which had good predictions for low and medium and k-NN was strong for low and Logistic Regression had a high accuracy for the high price.

**Introduction:**

Categorizing cameras into price ranges made the most sense to do and I broke it up into low which was anything less than $200. Medium was anything in between $200 and $600. Finally high was anything higher than $600. Model performance was analyzed using precision, recall, F1-score, and accuracy.

**Related work:**

Two similar pieces of work that I found that used machine learning, was this study on mobile price prediction. The first study (Majumder, 2022) had a lot more information than I did but they also had many more variables to consider. There is a lot more that does into pricing of phones as opposed to cameras. Nowadays it is not as common to see people purchasing cameras because everyone’s smart phone already comes equipped with a camera. The second study (ERCAN, 2023) was also similar to the first study which was also trying to predict phone prices.

**Mythology:**

My data came from Kaggle, and it was about cameras and their prices. There was a price category that I added to be able to separate the cameras based on their price. Low, medium, and high. There were also certain variables that were dropped to have cleaner data.

There were three models that were used for this project, logistic regression which is a linear model for multiclass classification. k-NN which is a distance-based model with 5 neighbors. Decision tree was a tree-based model with feature splits.

These values are going to be evaluated using the following metrics. Precision is the correct positive predictions. Recall is the proportion of actual positives that were correctly identified. The F1-score if the mean of precision and recall. Lastly, accuracy is the overall proportion of predictions.

**Results:**

Logistic regression was a linear model k-NN used 5 neighbors and went for the distance-based approach Decision tree splits into future thresholds#

Each model had their strengths and weaknesses. For example, linear regression had an accuracy of 59% for all predictions, but for high it had an 84%. On recall it was 86% accurate for low. On the F1-score low was .69 since it had high recalls, It struggled with medium recalls. Accuracy was low overall with 59%.

k-NN had an accuracy of 63% for all predictions. It had a 68% precision for. low. Recalls was a 78% on lows as well and a .73 on low F1-score score, making this model strong for low predictions. This made this model weak on high predictions.

Decision tree had an accuracy of 67% for all predictions. It had a 74% for low precision, 73% for low recall and a .73 for F1- score. This is good with low and medium overall the best one for predicting the prices.

Logistic Regression:

precision recall f1-score support

High 0.84 0.47 0.60 34

Low 0.57 0.86 0.69 99

Medium 0.55 0.29 0.38 75

accuracy 0.59 208

macro avg 0.65 0.54 0.56 208

weighted avg 0.61 0.59 0.56 208

k-NN:

precision recall f1-score support

High 0.49 0.56 0.52 34

Low 0.68 0.78 0.73 99

Medium 0.64 0.48 0.55 75

accuracy 0.63 208

macro avg 0.60 0.61 0.60 208

weighted avg 0.64 0.63 0.63 208

Decision Tree:

3

precision recall f1-score support

High 0.63 0.56 0.59 34

Low 0.74 0.73 0.73 99

Medium 0.59 0.64 0.62 75

accuracy 0.67 208

macro avg 0.66 0.64 0.65 208

weighted avg 0.67 0.67 0.67 208

**Conclusion:**

In this project an analysis was made to try and predict the price of certain cameras based on their properties utilizing the machine learning and classification There were three categories that were broken into the prices of low medium and high. The objective was to try and find a model that was going to have the highest accuracy. Predicting camera prices, the leading model was the decision tree. It has a 67% accuracy. The performance seemed to be the most balanced between the three groups.

Reference:

ERCAN, S. İ. A. (2023, May 12). *Mobile phone price classification using machine learning*. International Journal of Advanced Natural Sciences and Engineering Researches. https://as-proceeding.com/index.php/ijanser/article/view/791

Majumder, P. (2022, March 22). *Learn mobile price prediction through four classification algorithms*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2022/02/learn-mobile-price-prediction-through-four-classification-algorithms/