EE 500 Project
Luxury Handbag Price Prediction

Abinaya Manimaran, Ananya Singh, Deepika Kanade

Problem Motivation:

Most women love shopping, especially handbags. Luxury handbags can cost anywhere between $2000 to $40000. This project is to understand the pricing of these luxury handbags. We build various models from simple linear and ridge regression to complex multilayer perceptrons, convolutional neural networks, ensemble models with fuzzy approximation. We also compare the performances of all models and conclude the best performance regression for pricing.

Talking to Experts:

We decided to talk to four experts from the industry associated with the brands Celine, Gucci, Tom Ford and Louis Vuitton around Rodeo Drive, Los Angeles. Based on expert opinions, we came up with 27 features which could be considered for the price estimation of a luxury handbag. We also enquired about the major features which could potentially affect the price of a bag such as skin type.

Data Collection:

We collected 142 handbag data and a total of 27 features for every bag. The table below shows various features we considered for every bag:

Brand	Hardware – Metal Type	Number of Components	Usage Type
Bag Name	Hardware – Number of zips	Major Color	Number of Functionality
Accessories	Hardware - Strap type	Number of Colors	Launch Season
Bag Style	Hardware – strap length	Height, Width, Breadth	Launch Year
Skin Type	Number of Compartments	Volume	Production Type
Inner Material	Place of Manufacture	Demand	Number of Hours
Skin Source	PRICE		Number of People

Important Feature Identification:

Based on experts opinion, out of the all the features we considered to proceed, below are the features that are important in pricing luxury handbags.

- Brand - Chanel, Gucci, Celine, Hermes, Louis Vuitton, Tom ford
- Skin type - Crocodile, Alligator, Python, Snake, Lizard, Calf, Lamb
- Number of components in a bag - Single piece or different components stitched together
- Volume of the bag
- Production - Hand stitched or machine made bag

Pre-processing:

The features in the dataset comprised of categorical and numerical features. There features had to be preprocessed which included normalization and conversion of categorical features to numerical features. The data was split into train and test in a ratio of 80:20 leading to a training data of 113 points and a testing data of 29 points. The following preprocessing methods were performed:

- Normalization of continuous features: Number of zips, Strap length, Number of compartments, Number of components, Number of colors, Volume, Number of functionalities.
- Ordering Categorical features using Label Encoding - Metal type, Strap type, Inner material
- Ordering Categorical Features using One Hot Encoding - Brand, Accessories, Bag style, Major color, Skin Type

Initial Model With All Features:

We conducted initial analysis using all the features. Hyper-parameter tuning was performed in the case of regression with regularizers to get the least mean square error. Our observations are summarized in the table given below:

Model	**Test RMSE**	**Test R^2 Score**
Linear	4248.90	-0.367
Ridge (alpha = 3)	**2231.86**	**0.477**
Lasso(Alpha = 3)	2578.71	0.316
SVR(kernel = Poly)	3359.31	-0.323

We observed that using all the features in the analysis led to overfitting.

Multi-layer Perceptrons (MLP):

We tried to input the entire data into a Multi-layer perceptron model and tried various combinations of hidden layers, number of neurons in each layer and other hyper-parameters. The least RMSE value was obtained using the following combination:

- Loss – MSE
- Optimizer – Adam
- Learning Rate - 0.01
- Epochs – 1000

The least RMSE score obtained was 4159.08.

However, we concluded that the MLP was very unstable and was converging slowly.

Convolutional Neural Network (CNN):

The data collection for the CNN was done using web scraping where multiple views of a bag were downloaded from the brand website and resized and normalized to 32x32 to be the input. A total of 628 images were collected with each bag, at each price point had five views on an average. The train/test split was 80/20 with shuffling. Data Augmentation strategies included random zooming in and cropping with random horizontal flips. This was done to introduce invariance in the network.

The initial architecture chosen was the Lenet5 model (5 layers) and the final modified architecture consisted of a total of 6 layers with 3 convolutional layers of filter size 5x5 with 6, 16 and 6 filters respectively. The last two fully connected layers consisted of 120 and 84 neurons whereas the final layer consisted of an identity neuron with MSE as the loss function. This final neuron gives us the predicted price. The optimizer used is Adadelta with a learning rate of 0.001. The final results of the CNN architectures are given in a table below.

Model Architecture	**R^2 score**	**RMSE Loss**
Lenet, batch size = 64, #epochs = 1000	0.322	5052.8499
Modified architecture, batch size = 32, #epochs = 1000	0.357	4789.56
Modified architecture, batch size = 32, #epochs = 5000	0.511	3963.98
Lenet with data augmentation, batch size = 64, steps per epoch = 16, #epochs = 1000	0.14	5903.003
Modified architecture with data augmentation, batch size = 64, steps per epoch = 16, #epochs = 1000	0.1578	5892.3

We can see from the above table that data augmentation increases the RMSE loss. This could be due to the dataset included multiple views (including inside) were associated with a single price point. This caused the classifier to converge at a worse minimum when compared to the classifier trained without data augmentation. The model with the lowest RMSE loss and highest R^2 score had the modified Lenet architecture.

Fuzzy Approximation:

We had 7 numerical features in our data collected. We considered ASAM 2D code for fuzzy approximation. Since we can use only 2 features at a time for fuzzy approximation in our 2D model, we considered all 21 combinations of two features. We trained the model on our training data using different number of rules and figured out a total of 17 rules work the best for our purpose. We calculate the RMSE for test data for all 21 approximations and choose the best 3 combination with the lowest RMSE. They are given in the table below.

Best Combinations	**RMSE**
Number of Components, Volume	3100.421
Number of Components, Number of Colors	3119.442
Number of Colors, Volume	3062.386

Fuzzy Model With Regression (Ensemble Model):

We wanted to build an ensemble regression model. We combine the fuzzy approximations along with the categorical features and build a various regression models. The results are given in the table below. We choose the following features for this experiment.

- Categorical features - skin type and brand
- Numerical features - fuzzy approximations of the above shown three combinations of two features.

MODEL	**TEST RMSE**
Linear	2591.49
Ridge (alpha = 3)	1698.69
Lasso (alpha = 3)	2573.76
SVR (kernel = poly)	2197.56

We can see that from all the above mentioned models in the report, this ensemble model combining fuzzy approximations of numerical features and categorical features work the best with lowest test RMSE for price prediction.

Brand Specific Model:

We also wanted to experiment building individual pricing model for every brand. This is because pricing and the way some features are considered is different for different brands. Thus we build separate regression models for every brand and check their test RMSE. The RMSE shows that for some brands like Celine, Louis Vuitton and Gucci, the brand specific model works very well when compared to the general model.

Brand	Test RMSE
Hermes	6222.89
Celine	297.66
Tom Ford	1262.27
Chanel	1479.66
Louis Vuitton	795.00
Gucci	673.63

Challenges and Conclusion:

Some challenges in the project were:

- Less number of data leading to overfitting most of the times
- Variation in data across different brands
- Some data which the experts said highly influence pricing were not available
    - Production type
    - Number of hours to manufacture
    - Source of skin
    - Launch year and season

To conclude we really enjoyed doing this project and experimenting various models. We now have a solid understanding of luxury handbag pricing.