Machine learning for circular business models

Edvin Listo Zec & Olof Mogren

edvin.listo.zec@ri.se, olof.mogren@ri.se

Introduction

In this project we investigate the predictability of the end price of online auctions using machine learning techniques. The algorithms were trained and evaluated on data collected from a Swedish online auction site (www.tradera.se), and the data was split into training, validation, and test data. We hypothesize that the results from this task will be indicative of the predictability of the residual value of second-hand items on the market. Our hypothesis is that a similar technique will be useful to estimate the value of second-hand inventories, to help estimate the value of circular businesses. We have analysed the predictability of auction prices on a specific item category from the collected data set:” clothing”. The main result in this work investigates the feasibility to predict the auction end price given the text description and an image of an item.

Dataset

The whole dataset contains 88,511 items and was randomly split into training set of size 70,807 and a test set of size 17,704. Each item in the dataset has a starting price, an end price, an end date, number of bids, a title, a text description, and an image. The training set was further split into 67,267 (95%) training samples and 3541 (5%) validation samples. In figure 2 we see the histogram over the end prices for the training set for items below 1000 SEK, and note the long-tail distribution where most items are sold for small values, and very few items are expensive. Only 1% of items are above 1000 SEK. Figure 1 shows four example images from the dataset.

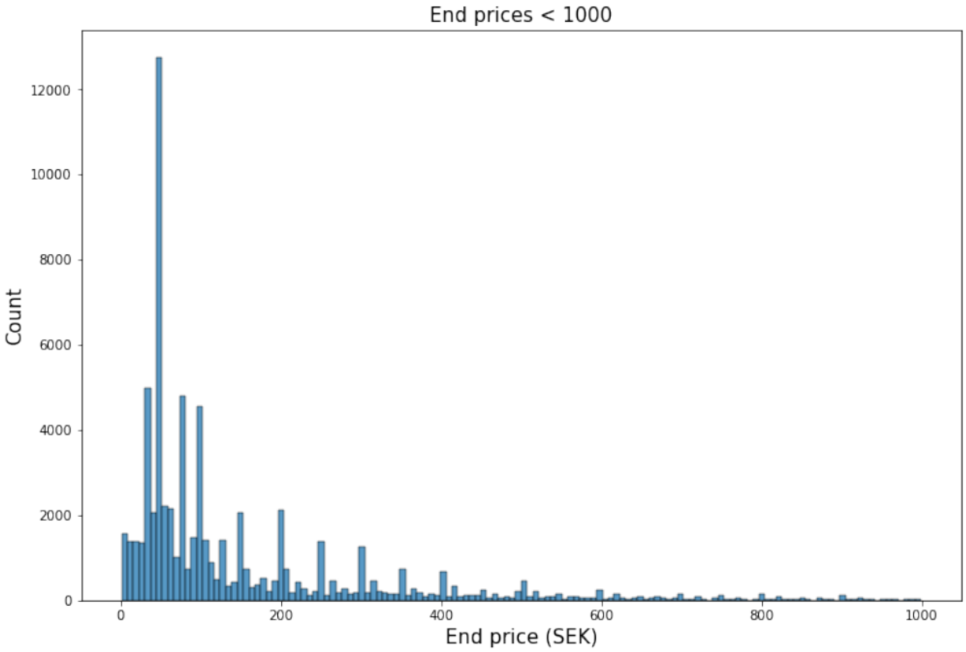

Figure 1: The end price distribution in the training set.

Figure 2: Example images from the training data

Input representations

In order to predict the auction end prices, we have used the title and the text description of each item as well as a user uploaded image of the item as input to the model. For the text descriptions, we have experimented with three different types of representations: unigrams, bigrams and Swedish CLIP embeddings. In figure 3 an illustration of our proposed method is shown.

CLIP is a large deep neural network that is trained using contrastive learning to match image and text pairs from a massive dataset of 400 million training samples. We use the pre-trained Swedish CLIP language model to create text representations of item descriptions, and we use the pre-trained CLIP vision model (a ResNet RN50x4) to create image representations. The text and image representations are both of dimension 640. These networks are only used to create representations, and are not updated during training.

The unigram and bigram representations were built from the training data, where all text was lower-cased and we set the maximum number of features to 10,000.

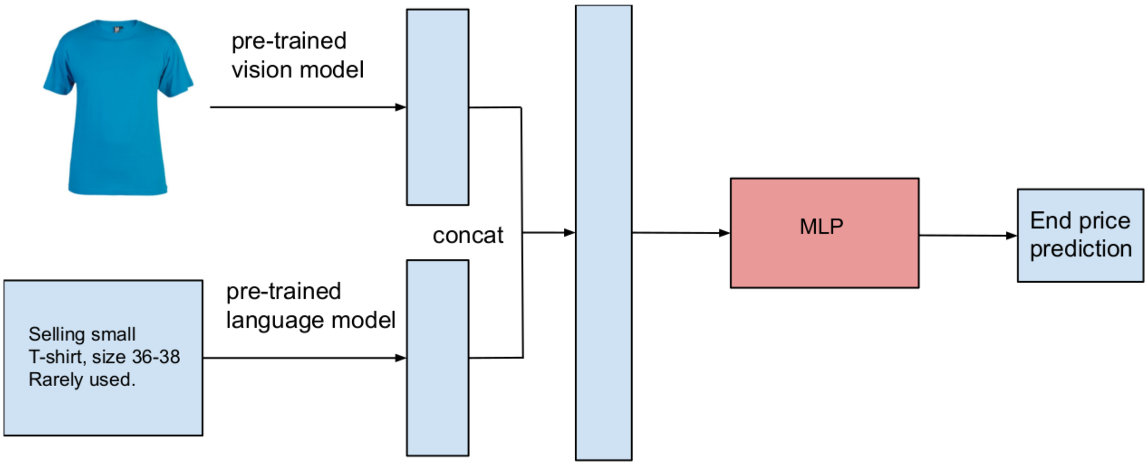

Figure 3: Schematic showing our proposed deep learning method

Model

In this work we are training two different models: a logistic regression model and a neural network. Logistic regression is a simple algorithm that can be trained to learn a mapping y = σ (Wx + b) from some input data x to classes y, where σ(x) is the sigmoid function. W is called a weight matrix and a bias term. During training, W and b are learned using the training data to minimize the error made by the model. A neural network can be viewed as an extension of this where we add layers of more weight matrices: y= W2(g(W1x + b)). This gives the model more capacity to learn more complex patterns in the data. g is called an activation function, and is a nonlinear function added to make it possible to learn non-linear patterns in the data. In figure 4 a visualization of a 1-layer neural network is shown.

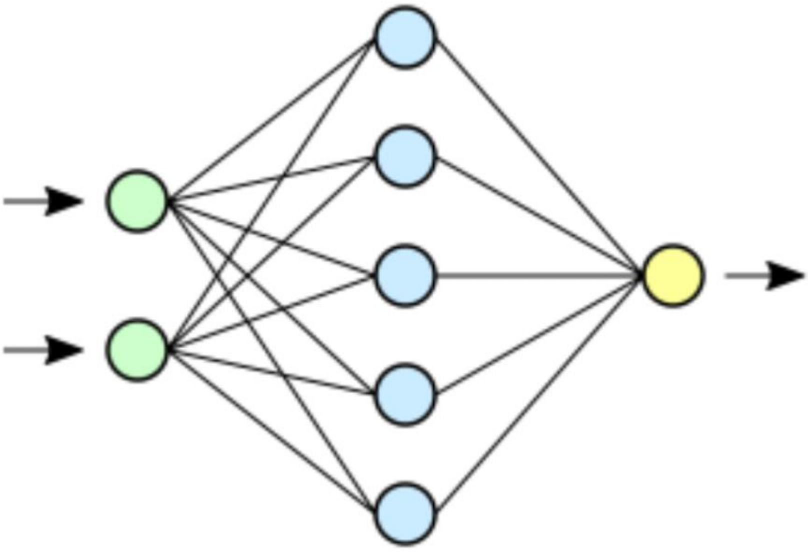

Figure 4: Visualization of a neural network with one hidden layer

Method

Our goal is to model the auction end prices. We compare the aforementioned representations together with two different models: logistic regression and a multi-layer perceptron (MLP). We will begin by treating the problem as a regression task. This will also be compared to classification task, since the long-tailed distribution of end-prices is hard to model. The task is then to predict which end price class each item falls into after an auction is finished. We investigate two different discretizations: four price classes and nine price classes. We manually create these classes such that their training distributions are as uniform as possible. See table 1 for the price classes.

Class	Price range (SEK)	
0	1-50	1-34
1	51-75	35-49
2	76-150	50
3	151+	51-79
4		80-103
5		104-154
6		155-249
7		250-400
8		400+

Table 1: Price class summary

Results

Classification

For the classification task, we evaluate the model performance using accuracy and confusion matrices. For the nine-class classification task, we also use top 2 accuracy, which is not used in the four-class setting due to the ranges becoming too wide and thus meaningless. The results are presented in tables 2 and 3. In both cases we see that the best performing model is using bigram text representation and clip image representations.

Model	Representation	Accuracy
MLP	Clip image	49.32
MLP	Clip text	53.18
MLP	Clip text+image	54.37
Logistic reg	Unigram	54.12
Logistic reg	Bigram	56.11
MLP	Bigram	57.03
MLP	**Bigram + clip image**	**57.40**

Table 2: Results for the four class classification models

Model	Representation	Accuracy	Top 2 accuracy
MLP	Clip image	30.25	47.56
MLP	Clip text	32.83	51.80
MLP	Clip text + image	33.86	53.06
Logistic reg	Unigram	34.33	53.46
Logistic reg	Bigram	36.08	56.03
MLP	Bigram	36.96	57.63
MLP	**Bigram + clip image**	**37.2**	**57.77**

Table 3: Results for the nine price class classification models

In figure 5, confusion matrices are shown for the four and nine-class tasks for the best performing model, i.e. the MLP using bigram and image representations. The values in the confusion matrices are normalized over the predictions, such that each column sums to 1.0. We note that the models have most of their errors close to the diagonal, meaning that when a model makes a prediction error, its prediction is close to the true label (not making large errors). This is further visualized in figure 6 where the error distribution (difference between true class and predicted class) is shown for the model in the nine-class task. We see here that the most common error value is 0, which means that the model made a correct prediction.

For financiers of circular business models it is of greater interest to assess the value of a stock of items, rather than a single item, as this helps more in assessing risks. We therefore evaluated if the classification model could estimate the value of a stock of items. To do that, we ran the best performing model through the 17,704 items in the test set and calculated the sum of the predicted price ranges. We discarded 1,773 items from class 8 (401+ SEK) since they did not have an upper range. The resulting estimate of the stock was $1,263,419- 1,928,556$ SEK for the model, and the true value of these items is 1,711,444 SEK, which lies in the predicted price range. Our results thus show that the model is also able to estimate the value of a stock of items to high precision.

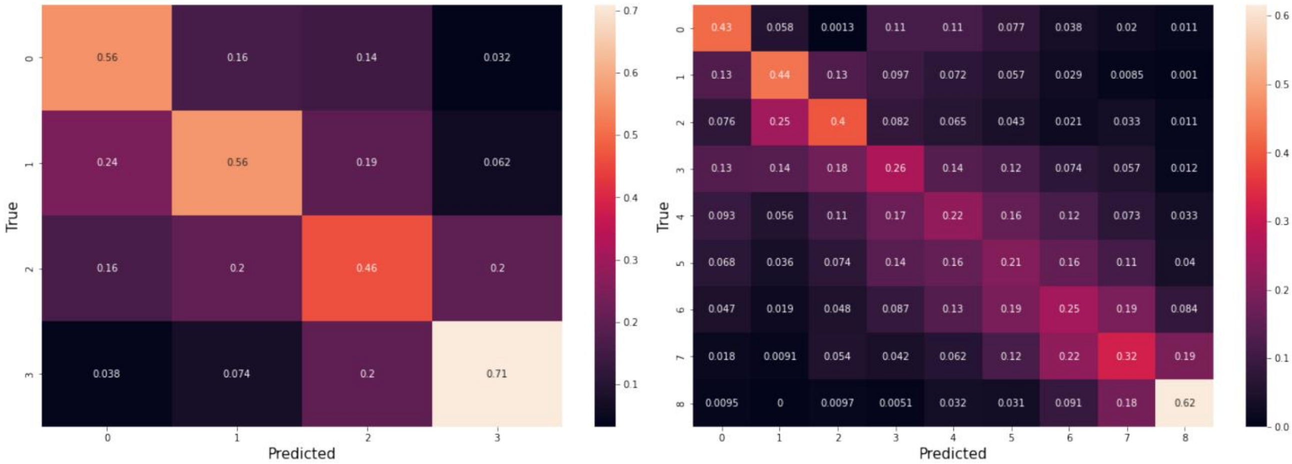

Figure 5: Confusion matrices for the classification problem (4 classes to the left and 9 classes to the right).

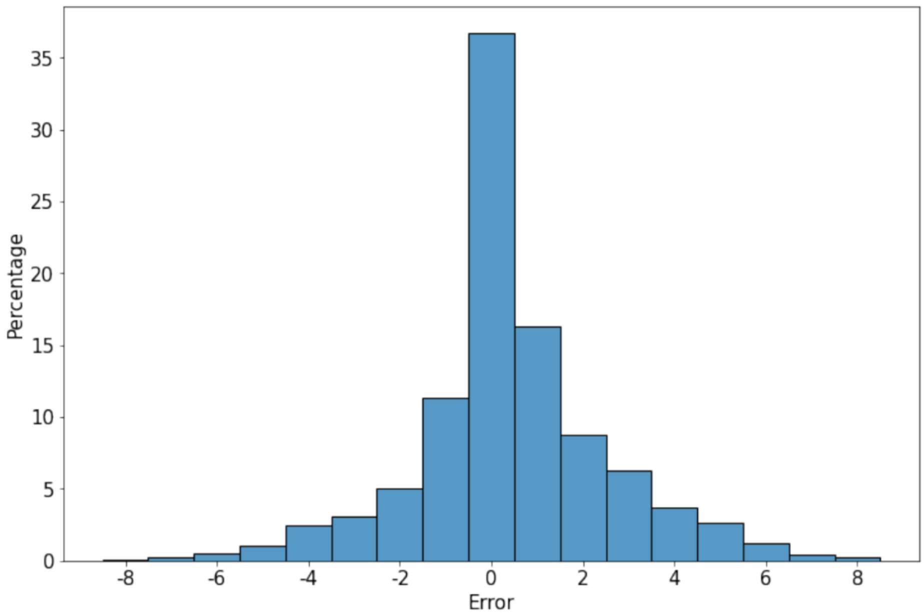

Figure 6

Regression

For the regression task, we trained a one hidden layer MLP to predict the end price for each item using bigram text representations and CLIP image representations as input. The MLP was trained using three different loss functions: mean absolute error (MAE), mean squared error (MSE) and mean squared logarithmic error (MSLE). All models were evaluated using the same three metrics, and we also compare them to a baseline which only guesses the median price from the training set. The results are summarized in table 4.

Model	MAE	MSE	MSLE	Stock value diff. ↓
Median guess	127.99	80,597	1.25	1,709,941 (54.70%)
MAE	**85.76**	40,661	**0.63**	549,016 (17.56%)
MSE	90.93	**35,196**	0.73	**46,970 (1.50%)**
MSLE	93.12	52,640	0.64	907,144 (29.02%)

Table 4

As we did for the classification task, we also compared how well the regression models captured the value of the whole test set (the sum of all end prices). We report the difference to the true test stock value in the last column of table 4. Using this metric, the model trained using the MSE loss outperforms all other ones. The total valuation of the test set is 3,126,261 SEK, and the model captures it with a 46,970 SEK error. This is 1.50% away from the true value.

Error distribution of the regression model is seen in figure 7.

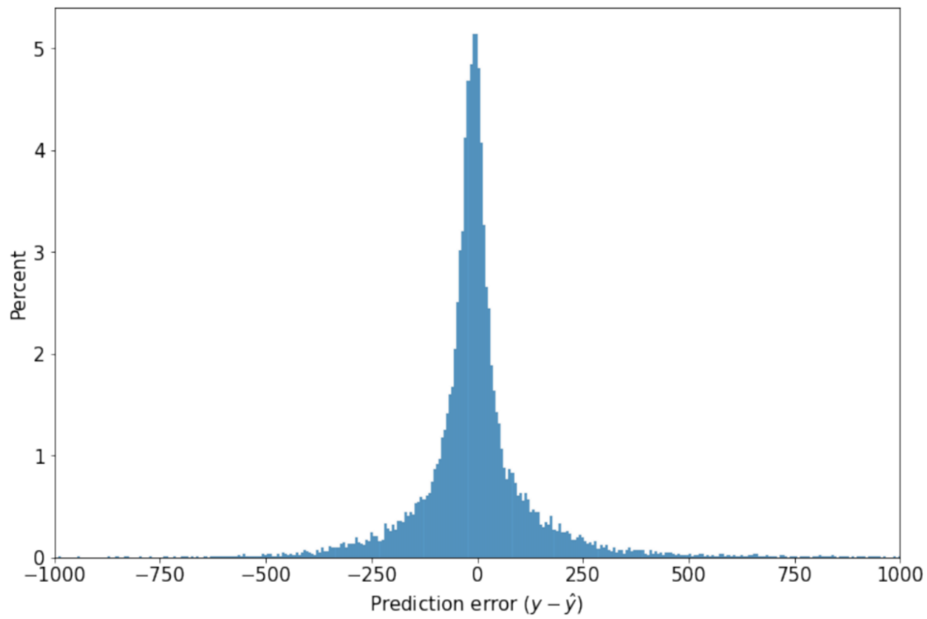

Figure 7: Prediction errors for the regression model on the test set

Human evaluation

As a further step of evaluation, we created a form consisting of 10 random images from the test set. This form was sent out to and answered by 32 humans, with the question asked being "*What do you think the end price of this auction was?".* The question was posed as a multiple-choice question, with the nine price classes as possible answers. The results for each human can be seen in figure 8 in blue. This can be compared to the deep learning model only using image representations as inputs (orange). The mean human accuracy was 18.75% on these 10 images, whereas the model achieved 40%. Only two humans got the same accuracy as the model, and only one human beat it. We also see that seven humans got a score of 0%. Further note that by using majority voting of the human answers an accuracy of 10%was achieved, not better than random chance. These results indicate how hard the problem is, and that the deep learning model is able to quite well estimate the value of second-hand items as compared to humans.

Human evaluation

As a further step of evaluation, we created a form consisting of 10 random images from the test set. This form was sent out to and answered by 32 humans, with the question asked being "*What do you think the end price of this auction was?".* The question was posed as a multiple-choice question, with the nine price classes as possible answers. The results for each human can be seen in figure 8 in blue. This can be compared to the deep learning model only using image representations as inputs (orange). The mean human accuracy was 18.75% on these 10 images, whereas the model achieved 40%. Only two humans got the same accuracy as the model, and only one human beat it. We also see that seven humans got a score of 0%. Further note that by using majority voting of the human answers an accuracy of 10% was achieved, *not better than random chance*. These results indicate how hard the problem is, and that the deep learning model is able to quite well estimate the value of second-hand items as compared to humans.

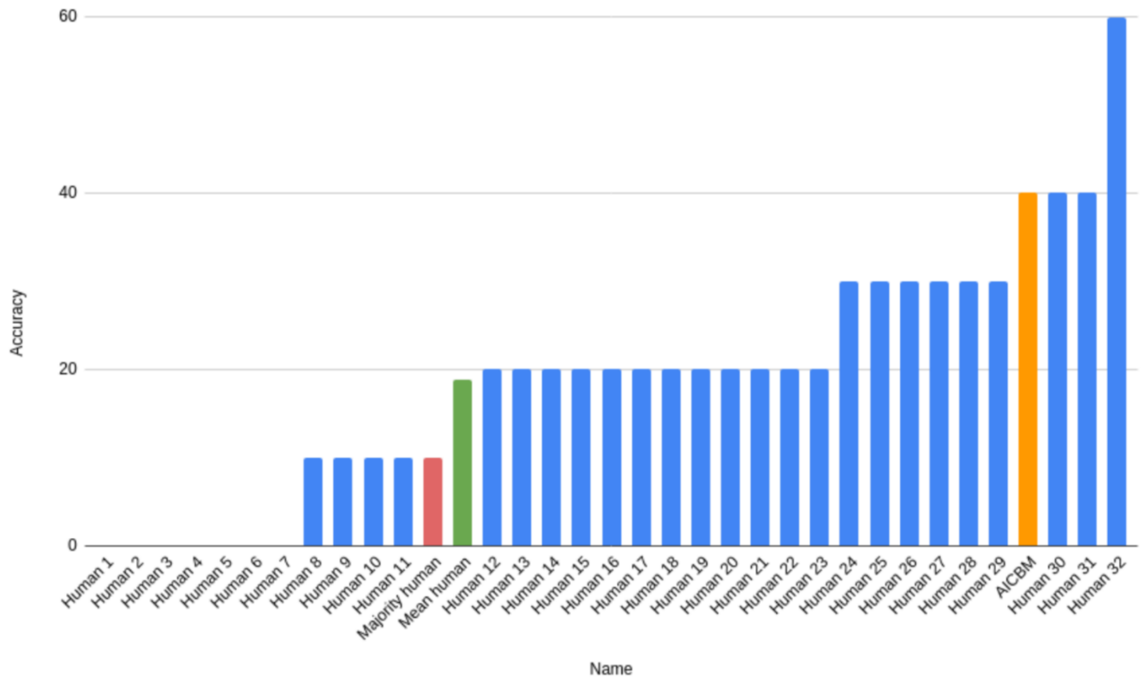

Figure 8: Accuracy of humans (blue), deep learning model (orange), mean human (green) and majority human vote (red).

Discussion

We have shown that it is possible to estimate auction end prices for the category of clothing. Our results show that image representations of auction items can be used to train a small neural network to model the residual value. Together with text representations from CLIP, the performance can be boosted. However, in the end the simplicity of only using unigram and bigram representations gave the best results. Our results also show that this is a hard problem even for humans to solve, and that our image AI model performed much better than average humans on the task of predicting price classes. For other future work, it would be interesting to collect and use more data in the modelling. If data is collected over several months, or even years, seasonality and trends could be used to further optimize when the best time to sell an item is, and estimate the profit. The collected data contains used items put online for sale by individuals. The advertisements contain misspellings, varying formatting, and photographs produced by amateurs without editing. This puts a cap on the achievable accuracy by a predictive model trained on the data. Further investigation should be put into looking at data that was more curated, or more uniformly produced. Such data may be available from online retailers such as Trove, who is running second-hand brand stores for brands such as Patagonia.