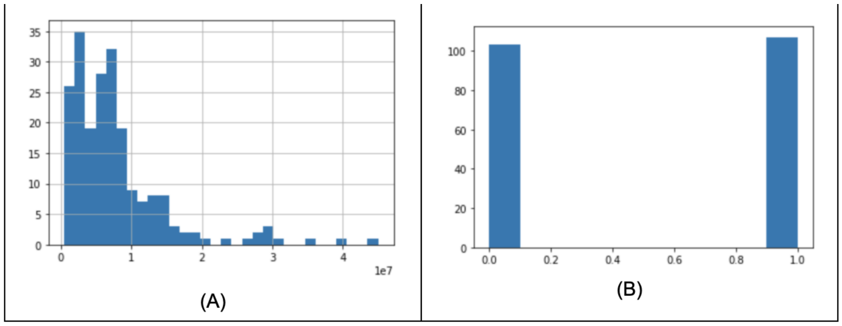

Figure 1: Depicted left, (A) presents the distribution of house prices in our dataset. On the right, (B) illustrates the roughly even distribution after dividing house prices into binary categories based on a price threshold.

begin with a logistic regression analysis to ground the experiments before introducing a deep MLP architecture. Our tabular data contains 5 features, categorical city and street names, bedrooms, and bathrooms, as well as continuous square feet. Our TF-IDF representation is sourced from a space-separated tokenization of words in the house descriptions. We do not include a special out of vocabulary token in our TF-IDF vectorization, such that if a word in the test set is not included in the TF-IDF vectorization formed from the training set, it is not represented during the test inference. Table 1 illustrates the performance of these models. The TF-IDF logistic regression achieves 2.4% higher accuracy than the tabular logistic regression. Further, TF-IDF logistic regression is able to perfectly fit the training data and still possess a reasonable generalization ability. This result inspired additional investigation into higher capacity models such as deep MLPs, as well as the use of dropout regularization to reduce the train-test generalization gap.

Deep Learning with Multi-Layer Perceptrons

Following experiments with logistic regression models, we explore the effectiveness of a deep MLP model for predicting house prices. Our MLP architecture is adapted for the two input sizes of 4873-d and 5-d for TF-IDF and tabular representations, respectively. Both data sources test the same MLP architecture of 3 hidden units containing 512, 1024, and 512 units. Each of these units are passed through a non-linear ReLU activation which prevents negative values in the intermediate activations. Each MLP makes a prediction by compressing the last layer of 512 units into a single prediction modulated with a sigmoid activation function. Due to the differences in input size, the TF-IDF MLP contains 5.5 million parameters, whereas the tabular MLP contains 1.05 million parameters. Table 1 presents the results of this experiment. The TF-IDF architecture perfectly fits to the training data, however, the logistic regression model performs at a 2.4% higher accuracy than the MLP model. In the following section, we turn to dropout regularization to close the performance gap between MLP and logistic regression.

The Impact of Dropout on MLP Predictions

Our initial experiment applying a Deep MLP architecture to TF-IDF representations of house descriptions achieved a perfect 100% training accuracy. However, this failed to generalize to the held-out test set, only achieving 76.2% test accuracy. Table 2 illustrates the benefit of applying dropout regularization to the MLP model. Dropout describes randomly zeroing out either inputs or intermediate activations, depending on where in the neural network architecture the dropout layer is placed. Dropout is used to regularize deep learning models and avoid overfitting such as what we have observed with 100% training accuracy and 76.2% testing accuracy. We utilize a high level of dropout at 75% placed at the input and in between every hidden layer of our 3-layer MLP network. This results in an increase in the test accuracy from 76.2% to 78.6% when processing TF-IDF data representations. Dropout similarly increases the tabular MLP from 71.4% to 73.8% test accuracy. Although dropout decreases the train-test generalization gap, we still have 100% training accuracy. We think this illustrates an opportunity for future improvement by exploring additional regularizations for deep neural networks, such as Data Augmentation (Shorten et al. 2021, Shorten and Khoshgoftaar 2019).

	Train Accuracy	Test Accuracy
Tabular Logistic Regression	79.3%	76.2%
Tabular MLP	87.5%	71.4%
TF-IDF Logistic Regression	100%	78.6%
TF-IDF MLP	100%	76.2%
BoW Logistic Regression	100%	47.6%
BoW MLP	85.1%	61.9%

Table 1: A comparison of tabular,TF-IDF, and BoW data inputs, as well as a comparison of logistic regression and MLP models.

Dropout %	0	75
TF-IDF Training Set	100%	100%
TF-IDF Testing Set	76.2%	78.6%
Tabular Training Set	87.5%	82.1%
Tabular Testing Set	71.4%	73.8%

Table 2: An illustration of the benefit of dropout regularization for deep MLP performance.