Machine Learning Project

We downloaded a dataset from Kaggle that described credit card fraud transactions. There were 28 features which described time of transaction cost of transaction and (whatever the rest meant). The data had approximately 30,000 training examples and we used 8500 examples for training, and 1500 for test and validation. The data was majority valid charges, with less than 1% being fraudulent charges. In our test set from 1500 examples \_\_\_ were fraudulent and the rest were valid. The approach we used was to read in the dataset into a 2d array and we used the holdout method and split the data into a train set, test set, and validation set. A support vector machine was trained on the data and had an accuracy of 100%.