Model Accuracy for Models Predicting an Increase in Price n Days from now Based on Past m days

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (m, n) | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 7) | (3, 3) | (4, 1) | (4, 2) | (7, 7) |
| Logistic Regression | 87.3% | 82.2% | 79.1% | 76.7% | 72.8% | 78.6% | 85.6% | 81.0% | 70.1% |
| Linear Discriminant Analysis | 87.3% | 82.2% | 79.1% | 76.6% | 72.8% | 78.6% | 85.6% | 81.0% | 70.1% |
| K Neighbors Classifier | 85.9% | 79.4% | 76.0% | 73.3% | 65.0% | 77.2% | 85.3% | 80.0% | 70.6% |
| Decision Tree Classifier | 80.9% | 74.4% | 70.8% | 68.8% | 65.5% | 72.0% | 79.7% | 74.9% | 66.3% |
| Gaussian NB | 87.3% | 82.2% | 79.1% | 24.5% | 72.7% | 28.7% | 16.5% | 26.9% | 65.8% |

Table 1: Table for model trained on all data points

Can’t use SVM because so many data points. Larger values of m do not lead to better accuracy, maybe only recent prices are useful to make predictions. Easier to predict the price tomorrow than for next week, long term predictions will be more inaccurate. (Solve this like that old website that was shut down? Inform users if price will fall 1, 3, 5, 7 days in the future. If all are increases, it’s probably safe to invest)

Model Accuracy for Models Predicting an Increase in Price n Days from now Based on Past m days

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (m, n) | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 7) | (3, 3) | (4, 1) | (4, 2) | (7, 7) |
| Logistic Regression | 86.4% | 81.1% | 77.8% | 75.5% | 71.48% | 77.5% | 84.9% | 79.9% | 69.1% |
| Linear Discriminant Analysis | 86.4% | 81.% | 77.8% | 75.5% | 71.5% | 77.4% | 84.9% | 79.9% | 69.2% |
| K Neighbors Classifier | 84.9% | 78.3% | 73.5% | 70.7% | 66.6% | 74.4% | 83.2% | 77.6% | 67.4% |
| Decision Tree Classifier | 78.8% | 72.5% | 68.8% | 67.5% | 63.6% | 69.5% | 77.4% | 72.2% | 63.3% |
| Gaussian NB | 63.8% | 55.1% | 48.5% | 53.7% | 47.6% | 48.4% | 38.4% | 43.3% | 44.4% |

Table 2: Table of model averages for one model per category

Model Accuracy for Models Predicting an Increase in Price n Days from now Based on Past m days

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (m, n) | (2, 1) | (2, 2) | (2, 3) | (2, 4) | (2, 7) | (3, 3) | (4, 1) | (4, 2) | (7, 7) |
| Logistic Regression | 82.3% | 75.3% | 71.1% | 69.0% | 64.8% | 70.7% | 78.3% | 73.0% | 59.9% |
| Linear Discriminant Analysis | 81.6% | 74.6% | 70.6% | 68.3% | 64.1% | 69.6% | 76.0% | 71.1% | 57.0% |
| K Neighbors Classifier | 81.3% | 73.8% | 70.1% | 67.5% | 63.3% | 69.2% | 78.0% | 71.9% | 59.2% |
| Decision Tree Classifier | 76.4% | 69.9% | 66.6% | 64.69% | 61.4% | 65.3% | 71.4% | 66.4% | 56.2% |
| Gaussian NB | 62.2% | 58.8% | 57.6% | 57.2% | 56.2% | 54.6% | 53.7% | 53.9% | 53.3% |

Table 3: Table of model averages for one model per item