**LITERATURE SURVEY**

1)**Predicting stock movement direction with machine learning**

**AUTHORS:** Jiao, Yang, and Jérémie Jakubowicz

Stocks movement direction forecasting has received a lot of attention. Indeed, being able to make accurate forecasts has strong implications on trading strategies. Surprisingly enough little has been published, relatively to the importance of the topic. In this paper, we reviewed how well four classic classification algorithms: random forest, gradient boosted trees, artificial neural network and logistic regression perform in predicting 463 stocks of the S&P 500. Several experiments were conduced to thoroughly study the predictability of these stocks. To validate each prediction algorithm, three schemes we compared: standard cross validation, sequential validation and single validation. As expected, we were not able to predict stocks future prices from their past. However, unexpectedly, we were able to show that taking into account recent information - such as recently closed European and Asian indexes - to predict S&P 500 can lead to a vast increase in predictability. Moreover, we also found out that, among various sectors, financial sector stocks are comparatively more easy to predict than others.

2)**Performance evaluation of predictive models for missing data imputation in weather data.**

**AUTHORS:**  Gad, Ibrahim, and B. R. Manjunatha

Real datasets can have missing values for a different reasons such as in data that were not kept on file and data corruption. Climate forecasting has a highly relevant effect in agricultural fields and industries sectors. The process of predicting climate conditions is required for different areas of life sectors. Handling missing data is significant because a lot of machine learning algorithms performance are affected by missing values in addition, they do not support data with missing values. Various techniques have been used to process missing data problem and the most applied is removing any row that contains at least one missing value. Also, another approaches to solve missing data problems are to impute the missing data to yield a more complete dataset. In order to improve the accuracy of prediction with the climate data, missing value from dataset should be removed or imputed/predicted in the pre-processing phase before using the data for prediction or clustering in the analysis step. In this paper, we propose a new technique to handle missing values in weather data using machine learning algorithms by execute experiments with NCDC dataset to evaluate the prediction error of five methods namely the kernel ridge, linear regression, random forest, SVM imputation and KNN imputation procedure. The missing values were imputed using each method and compared to the observed value. Results of the proposed method were compared with existing techniques

3)**Amazon EC2 Spot Price Prediction using Regression Random Forests**

**AUTHORS:**Khandelwal, Veena, Anand Chaturvedi, and Chandra Prakash Gupta

Spot instances were introduced by Amazon EC2 in December 2009 to sell its spare capacity through auction based market mechanism. Despite its extremely low prices, cloud spot market has low utilization. Spot pricing being dynamic, spot instances are prone to out-of bid failure. Bidding complexity is another reason why users today still fear using spot instances. This work aims to present Regression Random Forests (RRFs) model to predict one-week-ahead and one-day-ahead spot prices. The prediction would assist cloud users to plan in advance when to acquire spot instances, estimate execution costs, and also assist them in bid decision making to minimize execution costs and out-of-bid failure probability. Simulations with 12 months real Amazon EC2 spot history traces to forecast future spot prices show the effectiveness of the proposed technique. Comparison of RRFs based spot price forecasts with existing non-parametric machine learning models reveal that RRFs based forecast accuracy outperforms other models. We measure predictive accuracy using MAPE, MCPE, OOB Error and speed. Evaluation results show that MAPE <; = 10% for 66 to 92 percent and MCPE <; = 15% for 35 to 81 percent of one-day-ahead predictions with prediction time less than one second. MAPE <; = 15% for 71 to 96 percent of one-week-ahead predictions.

4)**On optimization methods for deep learning**

**AUTHORS:**Le, Quoc V., Jiquan Ngiam, Adam Coates, Abhik Lahiri, Bobby Prochnow, and Andrew Y. Ng

The predominant methodology in training deep learning advocates the use of stochastic gradient descent methods (SGDs). Despite its ease of implementation, SGDs are difficult to tune and parallelize. These problems make it challenging to develop, debug and scale up deep learning algorithms with SGDs. In this paper, we show that more sophisticated off-the-shelf optimization methods such as Limited memory BFGS (L-BFGS) and Conjugate gradient (CG) with line search can significantly simplify and speed up the process of pretraining deep algorithms. In our experiments, the difference between L-BFGS/CG and SGDs are more pronounced if we consider algorithmic extensions (e.g., sparsity regularization) and hardware extensions (e.g., GPUs or computer clusters). Our experiments with distributed optimization support the use of L-BFGS with locally connected networks and convolutional neural networks. Using L-BFGS, our convolutional network model achieves 0.69 % on the standard MNIST dataset. This is a state-of-theart result on MNIST among algorithms that do not use distortions or pretraining

5)**Semi-supervised learning literature survey**

**AUTHORS:**Zhu, Xiaojin

We review the literature on semi-supervised learning, which is an area in machine learning and more generally, artificial intelligence. There has been a whole spectrum of interesting ideas on how to learn from both labeled and unlabeled data, i.e. semi-supervised learning. This document is a chapter excerpt from the author’s doctoral thesis (Zhu, 2005). However the author plans to update the online version frequently to incorporate the latest development in the field.