**Resumos sobre os artigos e possíveis trabalhos relacionados**

(financial time series AND (forecasting OR prediction)) OR (Exchange Rate AND (Machine Learning OR prediction))

sentiment analysis AND (exchange rate OR stock market)

**Artigo 1**

Uma estratégia automatizada de investimento por meio de redes neurais artificiais e preditores econométricos **(estrategiaInvestimentoRedeNueralPreditor)**

**Algoritmos usados:**

Redes Neurais

SMA - Simple Moving Average

ARMA - Autoregressive Moving Average Model

ARIMA - Autoregressive Integrated Moving Average Model

GARCH - Generalized Autoregressive Conditional Heteroskedasticity

**Base de dados (Séries financeiras analisadas) usada:**

Companhia Energ´etica de Minas Gerais S.A. (CMIG4)

Embraer S.A. (EMBR3)

Itaú Unibanco Holding S.A. (ITUB4)

representando um índice foi utilizado o índice BOVESPA (BOVA11)

para um contrato futuro foi utilizada a commoditie de Boi Gordo (BGI$) negociada na Bolsa de Mercadorias & Futuros (BM&F) como derivativo

3 de janeiro de 2011 até 30 de dezembro de 2013

**Breve resumo**

a RNA com arquitetura Perceptron Múltiplas Camadas (PMC) + Preditores econométricos, foi estimado um conjunto de valores de preços de fechamento que serviu como entrada de um outra RNA PMC.

Foi comparado o desempenho da segunda RNA com os preditores econométricos, com a técnica buy and hold e com operações aleatórias.

linguagem de programação - R

plataforma para extração de dados - MetaTrader 5

Os resultados foram bons para os ativos do mercado de ações

Para o ativo de mercado futuro não foi bom, com lucros perto de zero

**Artigo 2**

Application of support vector machines in financial time series forecasting **(SVMinFinancialTimeSeriesStockIndexFuture)**

**Algoritmos usados:**

SVM com o estimador para problemas de regressão (basicamente a SVR)

Redes neurais

**Base de dados (Séries financeiras analisadas) usada:**

5 contratos futuros da Chicago Mercantile**,** que é uma bolsa de mercadorias dos EUA, baseada em chicago

Standard&Poor 500 stock index futures (CME-SP)

United States 30-year government bond (CBOT-US)

United States 10-year government bond (CBOT-BO)

German 10-year government bond (EUREX-BUND)

Frenchgovernment stock index futures (MATIF-CAC40)

CME-SP - 30/12/1992–30/07/1996

CBOT-US- 01/01/1993– 01/08/1996

CBOT-BO 01/01/1993– 01/08/1996

EUREX-BUND 01/01/1993– 01/08/1996

MATIF-CAC40 01/06/1995 – 01/02/1999

**Métricas usadas**

normalized mean squared error (NMSE)

mean absolute error (MAE)

directional symmetry (DS)

weighted directional symmetry (WDS)

**Breve resumo**

Para a SVM foi usado o kernel Gaussian

2 = 10 (parâmetro de kernel?) (símbolo de gama)

(um baixo valor ocasionou under-fitting, um alto valor over-fitting)

C = 100

e = 0.001

linguagem C++

Para saber os melhores valores dos parâmetros, ele pegou certos intervalos e foi incrementando aos poucos, os resultados estão plotados nos gráficos.

O preço de fechamento original é transformado em diferença relativa de cinco dias da porcentagem do preço (RDP)

As variáveis ​​de entrada são determinadas a partir de quatro valores RDP atrasados com base em períodos de 5 dias.

(fórmulas das variáveis de entrada e saída na tabela 2)

A vantagem mais proeminente é que a distribuição dos dados transformados se tornará mais simétrico e seguirá mais de perto uma distribuição normal, mostrado na **figura 1**.

data scaling é uma das técnicas de pré-processamento de dados usada

**Uma boa alternativa em relação as redes neurais, se saindo melhor do que ela em CME-SP, CBOT-US, CBOT-BO and MATIF-CAC40 e um pouco melhor em EUREX-BUND. Pois a SVM minimiza o limite superior do erro de generalização em vez de minimizar o erro de treinamento (isso de chama o princípio de minimização de risco), levando a uma melhor generalização.**

**Artigo 3**

Modeling and Trading the EUR/USD Exchange Rate Using Machine Learning Techniques **(EURUSDwithIA)**

**Algoritmos usados:**

K-Nearest Neighbors algorithm

Naïve Bayesian Classifier

Artificial Neural Networks

Support Vector Machines

Random Forest

Para a SVM foi usada a e Radial Basis Function (RBF) como função de kernel

C = 64

Gamma = 2

**Base de dados (Séries financeiras analisadas) usada:**

A taxa de câmbio entre EUR e USD no ECB (banco central europeu)

17 janeiro de 2002 até dezembro de 2010

**Métricas usadas**

Acurácia provavelmente....

**Breve resumo**

Apenas termos auto regressivos como entrada

Foi a realizado a **previsão da direção do movimento** de um dia a frente (problema de classificação)

Os algoritmos foram executados 10 vezes e tirado uma média

**transaction costs = custos de transação**

foram comparados com as estratégias tradicionais chamadas:

Naïve Strategy

moving average convergence / divergence technical model (MACD)

Random Forest foi o melhor, logo seguido da SVM nos termos de retorno anualizado e índice de avaliação / índice de informações (information ratio) mesmo quando os custos de transação (transaction costs) foram considerados.

No resultado com retorno anualizado (um tipo de cálculo de lucros levando em consideração o tempo para se obter tal lucro) e considerando os custos de transação o Random Forest teve o melhor resultado com 7.28% seguido logo atrás da SVM com 3.98%

KNN e o Naïve Bayesian foram os únicos que não superaram as estratégias tradicionais.

**Artigo 4**

**Foreign Exchange Trading with Support Vector Machines (****ForeignExhangeSVM)**

**Algoritmos usados:**

SVM usado como classificação, +1 para aumento, -1 para queda, ou seja, vai prever a direção da moeda.

C-Support Vector Classification (C-SVC)

Linguagem de programação R

**Base de dados (Séries financeiras analisadas) usada:**

As taxas de câmbio EUR/GBP, EUR/JPY e EUR/USD do período de 1 janeiro 1997 até 31 dezembro de 2004, totalizando 2349 dias de negociação.

**Procura achar uma correlação entre duas variáveis além dos limites de uma simples correlação, Casualidade de Granger (Granger causality) determina o sentido casual entre duas variáveis, estipulando que X “Granger causa” Y se os valores passados de X ajudam a prever o valor presente de Y.**

ele fala de algo até k=20

Causalidade de Granger (Granger causality) foi usada para determinar quais outras taxas tem fortes relações com as 3 de cima, foram pegos as com maiores influências para cada uma de cima. As escolhidas foram as utilizadas para serem as variáveis de entrada.

Por exemplo, EUR/GBP é Granger causada por 11 variáveis, entre elas: a taxa de câmbio EUR/USD, JPY/USD, os preços da platina e do níquel, os índices do mercado IBEX, MIB30, CAC e DJST, entre outras.

EUR/JPY é Granger causada por 11 variáveis, entre elas: a taxa de câmbio EUR/CHF, o índice de mercado IBEX

**Métricas usadas**

**Breve resumo**

um range de kernels foram comparados para saber qual dava um resultado melhor para a SVC, os kernels foram: Linear, Polinomial, Laplace, Gaussian, Hyperbolic e Bessel.

Para o benchmark são usados os modelo de precisão ingênuo (Naive) e o modelo econométrico ARMA

No geral o Hyperbolic foi o kernel superior

No caso da EUR/GBP os kernels Hyperbolic e Laplace foram os melhores, ficando bem iguais entre eles.

Estatisticamente a SVM foi superior.

**Artigo 5**

A Sentiment Analysis of Twitter Content as a Predictor of Exchange Rate Movements **(analisePrecoDolarComTwitter)**

**Algoritmos usados:**

**Base de dados (Séries financeiras analisadas) usada:**

A taxa de câmbio USD/TRY recuperados do Banco Central da República da Turquia (www.tcmb.gov.tr).

O período: 01.01.2013 – 31.12.2013

As palavras chaves usadas no Twitter foram: USD/TRY, #USD/TRY, Dollar, #Dollar

**Métricas usadas**

**Breve resumo**

Data mining e text mining são usados para prever mudanças na taxa de câmbio

Text mining serve para descobrir padrões dentro de um texto

Data mining serve para descobrir novos conhecimentos dentro de um grande conjunto de dados.

**Artigo 6**

Application of Support Vector Machine to Forex Monitoring **(SVMRegressionInForex)**

**Algoritmos usados:**

SVM regression model

**Base de dados (Séries financeiras analisadas) usada:**

**Breve resumo**

Forex se refere compra de uma moeda em pares, neste trabalho todas as modelas foram pareadas com o dólar australiano (AUD)

**Artigo 6**

**Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine (forecastingStockMarketMovementDirectionSVMSentiment.pdf)**

**Algoritmos usados:**

Máquina de vetor de suportes e HowNet and Chinese Sentiment Analysis

**Base de dados (Séries financeiras analisadas) usada:**

SSE 50 Index, é um índice do mercado da China, contem: preço de abertura, preço de fechamento, máxima do dia, mínima do dia, volume de negociação, volume de negociação em RBM, mudança em RBM e mudança em porcentagem.

Foi retirado os textos dos fóruns Sina stock ntro e Eastmoney stock fórum, do período de 17 de junho de 2014 até 7 de junho de 2016, incluindo 486 dias de negociação.

O numero total de comentários dos 2 foruns foi de 1.930.592, isso já depois de filtrar e retirar os ruídos do período determinado.

**Métricas usadas**

validação cruzada quíntupla (fivefold cross validation) e um protocolo de rolling window

**Breve resumo**

Criação de um modelo para auxiliar os investidores a tomarem decisões mais sábias, para prever a direção do movimento do preço (problema de classificação). Acurácia do previsão da direção do movimento foi de 89.93%, com um aumento de 18.6% depois da ntrodução de variáveis de sentimentos.

Foi utilizado uma estratégia de stop-loss para redução dos riscos.

Eles calcularam os sentimentos das 51 ações que representam o índice SSE 50 da China, porem perceberam que usar todos causa overfitting e decidiu usar apenar 8 desses 51

Foi implementado a validação cruzada quíntupla (fivefold cross validation), porem isso leva a um viés de antecipação (look-ahead bias), então integraram a SVM com um protocolo de rolling window para elimina-lo

Desenvolveram uma estratégia de negociação (trading) prática, 1 significa ordem de compra e -1 significa ordem de compra

sentiment indexes (índices de sentimentos) foram atribuídos ao modelo básico (botaram índices de sentimentos no algoritmo de SVM? Não entendi isso)

**Artigo 7**

**Forecasting the movement direction of exchange rate with polynomial smooth support vector machine**

**Algoritmos usados:**

support vector machine (PSSVM)

Broyden–Fletcher–Goldfarb–Shanno (BFGS)

**Base de dados (Séries financeiras analisadas) usada:**

O par de moeda RMB/USB, RMB é uma moeda da China

De 4 de janeiro de 2006 até 31 de outubro de 2006.

6 indices do Dow Jones China Index Series são usados como entrada

**Métricas usadas**

validação cruzada q

**Breve resumo**

Criação de um modelo

**Artigo 8**

**Forecasting stock market movement direction with support vector machine**

**Algoritmos usados:**

Support vector machine (SVM)

random walk model (RW)

Linear Discriminant Analysis,

Quadratic Discriminant Analysis

Elman Backpropagation Neural Networks.

**Base de dados (Séries financeiras analisadas) usada:**

Previsão da direção semanal do NIKKEI 225 index, esse índice compõe 225 ações da Bolsa de Valores de Tóquio.

O período da base de dados compõe de 1 de janeiro de 1990 até 31 de dezembro de 2002. Com um total de 676 pares de observações.

As variáveis macroeconômicas: S&P 500 Index e o par de moedas JPY/USD são escolhidas como entrada do modelo.

**Métricas usadas**

Ele usou os algoritmos: random walk model (RW), Linear Discriminant Analysis, Quadratic Discriminant Analysis e Elman Backpropagation Neural Networks para comparar com a SVM.

Então o desempenho relativo dos modelos é medido pela taxa de acerto (Hit Ratio)

**Breve resumo**

Compara o desempenho da SVM com os outros 4 algortimos, percebe que a SVM os supera. Depois propõe um modelo combinando todos os 5 algoritmos, e esta combinação obteve um resultado melhor que os outros obtiveram individualmente.

**Artigo 10**

**Support Vector Machine based Forecasting of the Contract Prices of Stock Index Futures**

**Algoritmos usados:**

support vector machines (SVM) para problemas de regressão.

**Base de dados (Séries financeiras analisadas) usada:**

Dados históricos de transações de Shanghai e o índice futuro de ações de Shenzhen na primeira temporada de IFSC3

O período foi de dezembro de 2008 até dezembro de 2009, um total de 238 registros de transações.

As variáveis de entrada do modelo são: preço de abertura, máxima, mínima, preço médio, volume de negociação, contratos em aberto e o preço de fechamento.

**Métricas usadas**

para calcular o resultado do modelo, foi usado uma formula de função de margem de erro mostrada em uma imagem no artigo.

**Breve resumo**

Foi usado para prever o preço de fechamento do dia seguinte, e depois comparado com o preço de fechamento do dia atual.

Nos testes, eles fizeram os testes em relação ao dia 20 de novembro, para a previsão do preço de fechamento do dia seguinte, o valor do dia seguinte era de 3999, e o modelo previu o preço de 3969.6, uma margem de erro de 0,74%.

**Artigo 11**

**Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting**

**Algoritmos usados:**

support vector machines (SVM) (The sequential minimal optimization (SMO))

Rede neural BP

Rede neural RBF

**Base de dados (Séries financeiras analisadas) usada:**

5 contratos futuros de Chicago Mercantile Market. São eles:

1 - CME-SP – de 04-01-1988 até 11-07-1995

2 - CBOT-US – de 01-10-1990 até 24-04-1998

3 - CBOT-BO – de 01-06-1989 até 11-12-1996

4 - EUREX-BUND - de 02-01-1991 até 06-08-1998

5 - MATIF-CAC40 – de 25-12-1991 até 30-06-1999

Os preços de fechamento diário são usados como conjunto de dados.

**Métricas usadas**

Mean squared error (NMSE)

Mean absolute error (MAE)

Directional symmetry (DS).

regularized RBF neural network também é usado para testes de banchmark

**Breve resumo**

A SVM foi melhor que a rede neural BP, e ficou equivalente a rede neural RBF. Porem a SVM com os parâmetros adaptativos teve resultados melhores que os outros modelos.

Sequential minimal optimization (SMO) é implementado neste trabalho para resolver o problema de regressão.

O kernel da SVM usado é a Gaussian function.

**Artigo 12**

**Social Media Aided Sentiment Analysis for Stock Prediction**

**Algoritmos usados:**

Maquina de vetor de suportes(SVM), J48 e o Naive Bayes são os algoritmos usados para a previsão das séries históricas.

**Base de dados (Séries financeiras analisadas) usada:**

As empresas escolhidas para se analisar o preço de fechamento delas são:

Infosys, Wipro, Hindustan Computer Limited (HCL), Tata Consultancy Services (TCS) e a Tech Mahindra junto com a BSE (SENSEX and Nifty), NSE e a Yahoo Finance

Preço do ouro(GP), taxa de cambio(ER) e o preço do petróleo bruto (COP).

O período dos dados é de janeiro de 2017 até julho de 2018. E são retiradas de fontes de dados como: NSE e BSE (duas bolsas de valores), Yaho finanças, controle de dinheiro e micro blog do twitter.

**Métricas usadas**

Accuracy

Sensitivity

Specificity

Root Mean Square Error (RMSE)

Mean Absolute Percentage Error (MAPE)

**Breve resumo**

Hence, the proposed method employs combined measures derived from technical, fundamental and sentiments. Twitter and Money Control act as a data source for providing opinion data to predict the stock price.

O artigo usa de previsão de séries históricas e analise de sentimentos para realizar as previsões da movimentação do preço dos ativos em questão.

No geral ele usa vários indicadores diferentes como entrada para a SVM, compara os resultados e percebe que a inclusão desse indicadores melhora o resultado final do algoritmo. Os indicadores são:

Indicadores da análise técnica:

RSI, Money Flow Index (MFI), Exponential Moving Average (EMA), Stochastic Oscillator (SO) e o Moving Average Convergence and Divergence (MACD)

Indicadores da análise fundamentalista:

Gold Price (GP), Exchange Rate (ER) and Crude Oil Price (COP) para o mercado de ações da bolsa de valores da Índia.

Indicadores da análise de sentimentos:

Aqui ele fez uma tabela com as palavras que indicam sentimentos positivos e as que indicam sentimentos negativos, extraídas dos textos do Money control e do Twitter.

**Artigo 12**

**Sentiment analysis on social media for stock movement prediction**

**Algoritmos usados:**

Support Vector Machine (SVM) voltado para problemas de classificação, com kernel linear

**Base de dados (Séries financeiras analisadas) usada:**

Duas bases de dados foram usadas, a primeira as séries históricas e a segunda é a base de dados de informações do humor.

As séries históricas foram recolhidas do yahoo Finance de 18 ativos diferentes. As informações recolhidas são: preço de abertura, máxima, mínima, fechamento e de fechamento ajustado.

Foi coletado 18 quadros de mensagens das 18 ações do Yahoo Finance Message Board.

O período da coleta de dados foi de 23 de julho de 2012 até 19 de julho de 2013.

**Métricas usadas**

Accuracy metric, a formula da acurácia usada foi mostrada no item 5.1, na poarte de Experiment Setup.

**Breve resumo**

A ideia é prever o valor do estoque em t, usando recursos derivativos em t-1 e t-2.

Foi analisado 18 ações em um ano. No geral o método final, unindo SVM e análise de sentimentos alcançou um desempenho 2.07% melhor que o modelo que usa apenas séries históricas. E comparando com apenas as ações mais difíceis de prever, o método teve 9,83% melhor precisão do que o método que usa apenas séries históricas, e 3,03% melhor que o método que usa apenas análise de sentimentos. Apesar do melhor resultado geral ter ficado em 54,41%, a ação AMZN chegou a 71,05%, ou a DELL que chegou a 64,47%.

No artigo é proposto um ‘topic-sentiment’, onde a ideia é descobrir quais tópicos sobre o mercado de ações as pessoas estão falando mais nas redes sociais, como por exemplo o sentimento sobre os tópicos específicos daquela empresa, o produto, serviço, dividendo e assim por diante.

No geral o artigo criou 6 modelos diferentes a partir da SVM, variando entre, alguns usam apenas séries históricas, outros apenas os sentimentos, outros a combinação dos 2 (tudo isso mostrado na tabela 3), depois ele criou uma tabela para comparar o resultados dos 6 modelos na previsão das 18 ações diferentes (tabela 6), e mostrou a taxa de acerto (acurácia) em cada ação.

**Artigo 13**

**Financial time series forecasting using support vector machines**

**Algoritmos usados:**

Support Vector Machine (SVM) voltado para problemas de classificação

Back-propagation neural networks (BP) Case-based reasoning (CBR) (estes usados para comparar com o resultado da SVM).

**Base de dados (Séries financeiras analisadas) usada:**

Korea Composite Stock Price Index (KOSPI) , é índice da bolsa de valores da Coreia do Sul

O período coletado foi de janeiro de 1989 até Dezembro de 1998, totalizando 2928 dias de amostra.

**Métricas usadas**

A métrica usada foi a Hit ratio.

**Breve resumo**

O artigo foca em problemas de classificação, para prever a direção de mudança diária do índice KOSPI. O numero ‘0’ indica que o preço do dia seguinte é menor que o do dia atual e ‘1’ indica que o preço do dia seguinte é maior que o do dia atual.

O artigo selecionou um total de 12 indicadores da análise técnica, para servir como atributos iniciais para o algoritmo. Eles estão listados na Tabela 1.

Os kernels pra a SVM usados foram: polynomial kernel e o Gaussian Kernel.

Nos resultados a SVM se saiu melhor que a BP e a CBR, o resultado final indicou uma acurácia de 57,83% para a SVM, comparado com os 54,73% da BP e 51,97% da CBR.