Data Science for Retail

Retail Analytics using Advanced Machine Learning

AlgoAnalytics

Outline

About **AlgoAnalytics**

Analytics for **Offline** Retail

AlgoAnalytics

CEO and Company Profile

About AlgoAnalytics

Analytics Consultancy

- Work at the intersection of mathematics and other domains
- Harness data to provide insight and solutions to our clients

Led by Aniruddha Pant

- +30 data scientists with experience in mathematics and engineering
- Team strengths include ability to deal with structured/ unstructured data, classical ML as well as deep learning using cutting edge methodologies

Expertise in Mathematics and Computer Science

- Develop advanced mathematical models or solutions for a wide range of industries:
- Financial services, Retail, economics, healthcare, BFSI, telecom, …

Working with Domain Specialists

- Work closely with domain experts – either from the clients side or our own – to effectively model the problem to be solved

Aniruddha Pant

CEO and Founder of AlgoAnalytics

PhD, Control systems, University of California at Berkeley, USA 2001

Highlights

- 20+ years in application of advanced mathematical techniques to academic and enterprise problems.
- Experience in application of machine learning to various business problems.
- Experience in financial markets trading; Indian as well as global markets.

Expertise

- Experience in cross-domain application of **basic scientific process**.
- Research in areas ranging **from biology to financial markets to military applications**.
- Close collaboration with premier educational institutes in India, USA & Europe.
- Active involvement in startup ecosystem in India.

Prior Experience

- Vice President, Capital Metrics and Risk Solutions
- Head of Analytics Competency Center, Persistent Systems
- Scientist and Group Leader, Tata Consultancy Services

AlgoAnalytics

AlgoAnalytics - One Stop AI Shop

Financial Services

- Dormancy prediction
- Recommender system
- News summarization – automated 60 words news summary

Healthcare

- Medical Image Diagnostics
- Work flow optimization
- Cash flow forecasting

Legal

- Contracts Management
- Structured Document decomposition
- Document similarity in text analytics

Internet of Things

- Assisted Living
- Predictive in ovens
- Air leakage detection
- Engine/compressor fault detection

Others

- Algorithmic trading strategies
- Risk sensing – network theory
- Network failure model
- Multilanguage sentiment analytics

- We use structured data to design our predictive analytics solutions like churn, recommender sys
- We use techniques like clustering, Recurrent Neural Networks,

Structured Data

- We used text data analytics for designing solutions like sentiment analysis, news summarization and many more
- We use techniques like natural language processing, word2vec, deep learning, TF-IDF

Text Data

- Image data is used for predicting existence of particular pathology, image recognition and many others
- We use techniques like deep learning – convolutional neural network, artificial neural networks and technologies like TensorFlow

Image Data

- We use sound data to design factory solutions like air leakage detection, identification of empty and loaded strokes from press data, engine-compressor fault detection
- We use techniques like deep learning

Sound Data

AlgoAnalytics

Technologies

NLTK

R

TensorFlow

theano

Pandas

python

Microsoft Azure

cassandra

mongoDB

php

PostgreSQL

C++

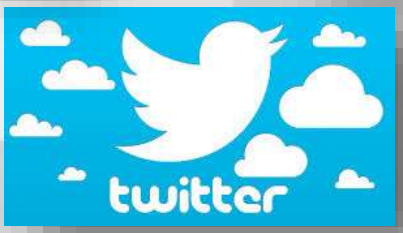
Microsoft CNTK

Java

Spark

Google news

H_2O.ai



AlgoAnalytics

Analytics in Offline Retail

Price Optimization

- Dynamic pricing based on demand and profit margins
- Devising offers strategically leading to increased revenue

Supply chain logistics

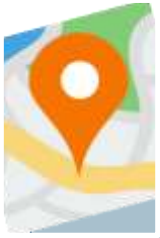
- Reduced logistic costs to prevent revenue leakages
- Returns prediction for efficient logistic solutions

Sales Forecasting

- Plan ahead and modify policies for higher sales
- Strategize marketing campaigns based on the forecasts

Inventory Forecasting

- Higher returns on the capital with accurate inventory forecasts
- Optimized warehouse stocks and better demand-supply management



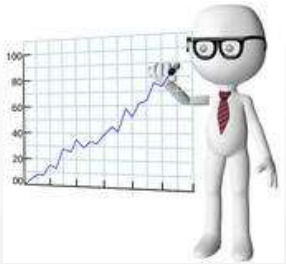
Location Analytics

- Leveraging demographic data (age, education, income, preferences, etc.) for sales improvement
- Planning potentially best locations for new stores

AlgoAnalytics

Sales Forecasting

- A **time-series** is a dataset that has values over a period of time.
- Sales Forecasting is future prediction for sales based on past sales performance (time-series)



Analyse sales and Forecast

Plan ahead looking at the forecast

Higher profits with better planning

Why Forecast Sales?

Enables viewing the future objectively	Using the forecasts one can establish policies to monitor prices and other costs	Manufacturing industries can plan for production and capacity	Retail companies can form basis for marketing

AlgoAnalytics

Pre-Forecasting Data Analysis

- Seasonality in the data
  - Seasonal patterns refers to a fixed period influencing sales like holiday season or a particular month or weekday
- Year over year trends
  - Analyze each years‘ worth of data separately to look at the trends
- Correlation of lags
  - How is the target’s sales dependent on the previous sales
- External factors affecting the sales, like offers, weather, etc.

Daily, Weekly and Monthly Features

Holidays’ impact over sales

Weekends generally see higher sales

Promos and offers

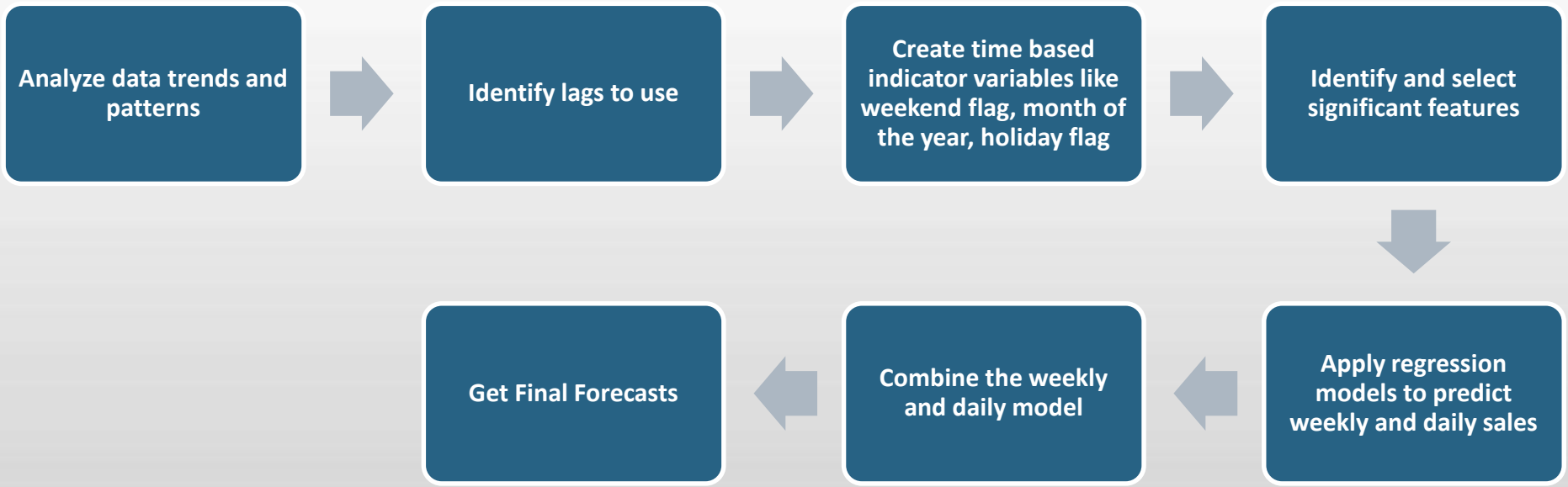
Geo location of the store and demographics

AlgoAnalytics

Sales Forecast: Retail Store Chain

Problem Definition: Forecast sales for each of the 45 days in future for all stores (~1200) in the chain using the daily sales data for last 3 years.
Dataset: Three major data columns, Date, Store ID and Sales in USD.

Steps followed for each store:



AlgoAnalytics

Sales Forecast: Retail Store Chain

Actual Sale	Predicted Sale	Error
$121,325	$123,674	-2349
$154,923	$154,784	139
$85,848	$84,475	1373
…	…	…

$$\text{RMSE} = \sqrt{\frac{(-2349)^2+(139)^2+(1373)^2+\cdots}{\textit{Total Observations}}}$$

As the name suggests, Root Mean Squared Error is the square root of average squared error

$$\text{MAPE} = \frac{\left(\frac{|-2349|}{123674}\right)+\left(\frac{|139|}{154784}\right)+\left(\frac{|1373|}{84475}\right)+\cdots}{\textit{Total Observations}}$$

Mean Absolute Percentage Error is the mean of absolute error percentage on the sales.

Results

RMSE

- Root Mean Squared Error
- Frequently used metric for calculating error in predictions
- Observed RMSE = **~4200**

MAPE

- Mean Absolute Percentage Error
- Measure of prediction accuracy of a forecasting method in statistics
- MAPE = **~18%**

AlgoAnalytics

Sales Forecast: Manufacturing Industry

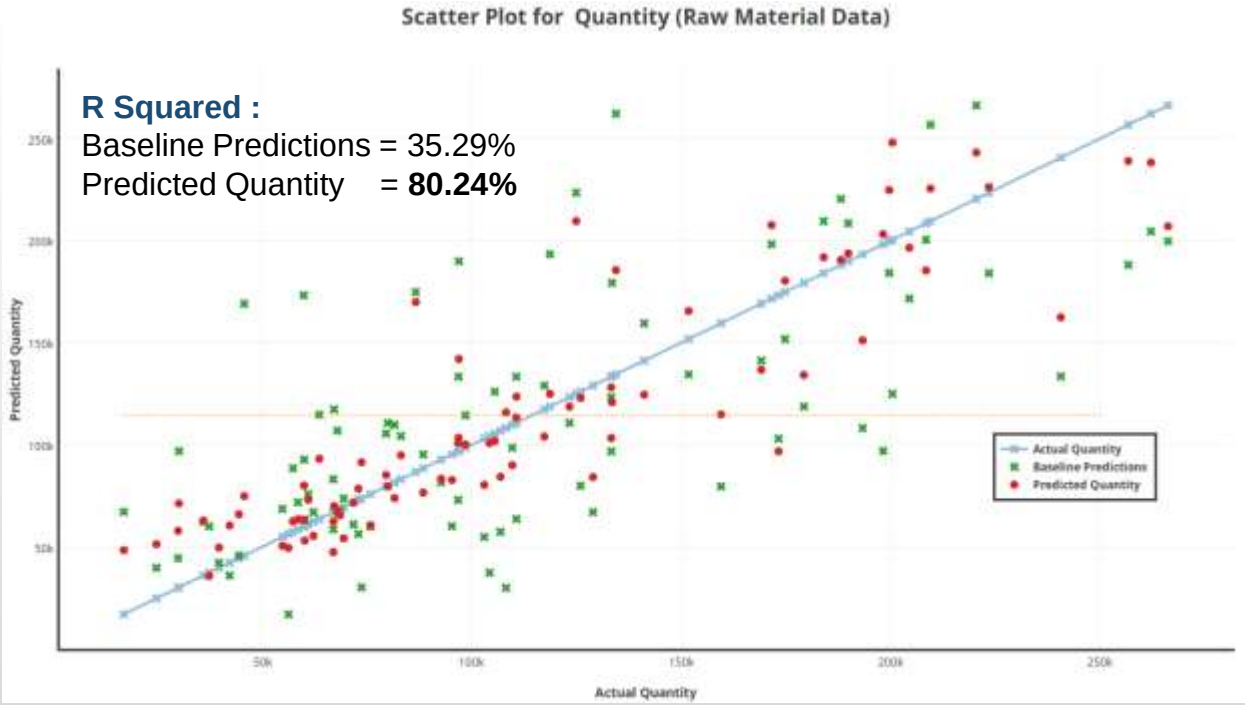
Problem Definition: Forecasting sales for the next month; for all the parts manufactured and the raw materials needed, using the monthly sales and cost data for last 45 months.

Normalize sales/quantity matrix for all months and parts

Apply PCA for dimensionality reduction

Find top 3 similar months using nearest neighbors (KNN)

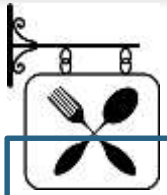
Weighted average for next month to the top 3 neighbors (KNN)



The R Squared values suggest that on an average the predictions are over 2 times better (closer to the actuals) than the naive predictions

AlgoAnalytics

Sales Forecast: Case Studies



Kaggle Competitions

ROSSMANN

Restaurant Revenue Prediction

- Predict the sales for newly opened restaurants given the data for older restaurants
- Store location city and some precomputed features for real estate, demographics etc. are given
- We used Random Forest algorithm for predictions as against Gradient Boosting Trees by the competition winner
- Final post deadline submission was **better than the competition winner!**

Rossmann Store Sales

- Forecast sales for 3000 Rossmann stores for up to 6 weeks in advance
- School holiday, Competitor store distance, promos, etc. features are provided
- Fourier transform on features was used on some of the variables
- **Final rank (post deadline) achieved was 74** using an ensemble of linear regression and XG boost model.

AlgoAnalytics

Interested in knowing more:

Contact us: info@algoanalytics.com

AlgoAnalytics