

Business Machine Learning and Data Science Applications

A curated list of applied business machine learning (BML) and business data science (BDS) examples and libraries. The code in this repository is in Python (primarily using jupyter notebooks) unless otherwise stated. The catalogue is inspired by awesome-machine-learning.

Table of Contents

Department Applications

- **Accounting**
 - Machine Learning
 - Analytics
 - Textual Analysis
 - Data
 - Research and Articles
 - Websites
 - Courses
- **Customer**
 - Lifetime Value
 - Segmentation
 - Behaviour
 - Recommender
 - Churn Prediction
 - Sentiment
- **Employee**
 - Management
 - Performance
 - Turnover
 - Conversations
 - Physical
- **Legal**
 - Tools
 - Policy and Regulatory
 - Judicial
- **Management**
 - Strategy
 - Decision Optimisation

- Causal Inference
 - Statistics
 - Quantitative
 - Data
- **Operations**
 - Failures and Anomalies
 - Load and Capacity Management
 - Prediction Management

Accounting

Machine Learning

- Chart of Account Prediction - Using labeled data to suggest the account name for every transaction.
- Accounting Anomalies - Using deep-learning frameworks to identify accounting anomalies.
- Financial Statement Anomalies - Detecting anomalies before filing, using R.
- Useful Life Prediction - Predict the useful life of assets using sensor observations and feature engineering.
- AI Applied to XBRL - Standardized representation of XBRL into AI and Machine learning.

Analytics

- Forensic Accounting - Collection of case studies on forensic accounting using data analysis. On the lookout for more data to practice forensic accounting, *please get in touch*
- General Ledger - Data processing over a general ledger as exported through an accounting system.
- Bullet Graph - Bullet graph visualisation helpful for tracking sales, commission, and other performance.
- Aged Debtors - Example analysis to investigate aged debtors.
- Automated FS XBRL - XML Language, however, possibly port analysis into Python.

Textual Analysis

- Financial Sentiment Analysis - Sentiment, distance, and proportion analysis for trading signals.
- Extensive NLP - Comprehensive NLP techniques for accounting research.

Data, Parsing and APIs

- EDGAR - A walk-through in how to obtain EDGAR data.
- IRS - Accessing and parsing IRS filings.
- Financial Corporate - Rutgers corporate financial datasets.
- Non-financial Corporate - Rutgers non-financial corporate dataset.
- PDF Parsing - Extracting useful data from PDF documents.
- PDF Tabel to Excel - How to output an excel file from a PDF.

Research And Articles

- Understanding Accounting Analytics - An article that tackles the importance of accounting analytics.
- VLFeat - VLFeat is an open and portable library of computer vision algorithms, which has Matlab toolbox.

Websites

- Rutgers Raw - Good digital accounting research from Rutgers.

Courses

- Computer Augmented Accounting - A video series from Rutgers University looking at the use of computation to improve accounting.
- Accounting in a Digital Era - Another series by Rutgers investigating the effects the digital age will have on accounting.

Customer

Lifetime Value

- Pareto/NBD Model - Calculate the CLV using a Pareto/NBD model.
- Cohort Analysis - Cohort analysis to group customers into mutually exclusive cohorts measured over time.

Segmentation

- E-commerce - E-commerce customer segmentation.
- Groceries - Segmentation for grocery customers.
- Online Retailer - Online retailer segmentation.
- Bank - Bank customer segmentation.
- Wholesale - Clustering of wholesale customers.
- Various - Multiple types of segmentation and clustering techniques.

Behaviour

- RNN - Investigating customer behaviour over time with sequential analysis using an RNN model.
- Neural Net - Demand forecasting using artificial neural networks.
- Temporal Analytics - Investigating customer temporal regularities.
- POS Analytics - Analytics driven customer behaviour ranking for retail promotions using POS data.
- Wholesale Customer - Wholesale customer exploratory data analysis.
- RFM - Doing a RFM (recency, frequency, monetary) analysis.
- Returns Behaviour - Predicting total returns and fraudulent returns.
- Visits - Predicting which day of week a customer will visit.
- Bank: Next Purchase - A project to predict bank customers' most probable next purchase.
- Bank: Customer Prediction - Predicting Target customers who will subscribe the new policy of the bank.

- Next Purchase - Predict a customers' next purchase also using feature engineering.
- Customer Purchase Repeats - Using the lifetimes python library and real jewellery retailer data analyse customer repeat purchases.
- AB Testing - Find the best KPI and do A/B testing.
- Customer Survey - Example of parsing and analysing a customer survey.
- Happiness - Analysing customer happiness from hotel stays using reviews.
- Miscellaneous Customer Analytics - Various tools and techniques for customer analysis.

Recommender

- Recommendation - Recommend the songs that a customer on a music app would prefer listening to.
- General Recommender - Identifying which products to recommend to which customers.
- Collaborative Filtering - Customer recommendation using collaborative filtering.
- Up-selling - Analysis to identify up-selling opportunities.

Churn Prediction

- Ride Sharing - Identify customer churn rates in order to target customers for retention campaigns.
- KKDBox I - Variational deep autoencoder to predict churn customer
- KKDBox II - A three step customer churn prediction framework using feature engineering.
- Personal Finance - Predict customer subscription churn for a personal finance business.
- ANN - Churn analysis using artificial neural networks.
- Bike - Customer bike churn analysis.
- Cost Sensitive - Cost sensitive churn analysis driven by economic performance.

Sentiment

- Topic Modelling - Topic modelling on a corpus of customer surveys from the VR industry.
- Customer Satisfaction - Predict customer satisfaction using Kaggle data.

Employee

Management

- Personality Prediction - Predict Big 5 Personality from text.
- Salary Prediction Resume - Textual analyses over resume to predict appropriate salary [Project Disappeared, still a cool idea]
- Employee Review Analysis - Review analytics for top 50 retail companies on Indeed.
- Diversity Analysis - A simple analysis of gender and race disparity in the tech industry.
- Occupation Prediction - Predict the likelihood that an occupation is analytical.

Performance

- Training Hours Performance - The impact of training ours on employee performance.

- Promotion Prediction - Analysing promotion patterns.
- Employee Attendance prediction - Various tools to predict employee attendance.

Turnover

- Early Leaving Employees - Identifying why the best and most experienced employees leaving prematurely.
- Employee Turnover - Identifying factors associated with employee turnover.

Conversations

- Slack Communication Analysis - Producing meaningful visualisations from slack conversations.
- Employee Relationships from Conversations - Identifying employee relationships from emails for improved HR analytics.
- Categorise Employee Requests - Classifying employee requests via TFDIF Vectorizer and RandomForestClassifier.

Physical

- Employee Face Recognition - A face recognition implementation.
- Attendance Management System - An attendance management system using face recognition.

Legal

Tools

- LexPredict - Software package and library.
- AI Para-legal - Lobe is the world's first AI paralegal.
- Legal Entity Detection - NER For Legal Documents.
- Legal Case Summarisation - Implementation of different summarisation algorithms applied to legal case judgements.
- Legal Documents Google Scholar - Using Google scholar to extract cases programmatically.
- Chat Bot - Chat-bot and email notifications.

Policy and Regulatory

- GDPR scores - Predicting GDPR Scores for Legal Documents.
- Driving Factors FINRA - Identify the driving factors that influence the FINRA arbitration decisions.
- Securities Bias Correction - Bias-Corrected Estimation of Price Impact in Securities Litigation.
- Public Firm to Legal Decision - Embed public firms based on their reaction to legal decisions.

Judicial Applied

- Supreme Court Prediction - Predicting the ideological direction of Supreme Court decisions: ensemble vs. unified case-based model.

- Supreme Court Topic Modeling - Multiple steps necessary to implement topic modeling on supreme court decisions.
- Judge Opinion - Using text mining and machine learning to analyze judges' opinions for a particular concern.
- ML Law Matching - A machine learning law match maker.
- Bert Multi-label Classification - Fine Grained Sentiment Analysis from AI.
- Some Computational AI Course - Video series Law MIT.

Management

Strategy

- Topic Model Reviews - Amazon reviews for product development.
- Patents - Forecasting strategy using patents.
- Networks - Business categories from Yelp reviews using networks can help to identify pockets of demand.
- Company Clustering - Hierarchical clusters and topics from companies by extracting information from their descriptions on their websites
- Marketing Management - Programmatic marketing management.

Decision Optimisation

- Constraint Learning - Machine learning that takes into account constraints.
- Fairlearn - I think it is called cost-sensitive machine learning.
- Multi-label Classification - Cost-Sensitive Multi-Label Classification
- Multi-class Classification - Cost-sensitive multi-class classification (Weighted-All-Pairs, Filter-Tree & others)
- CostCla - Costcla is a Python module for cost-sensitive machine learning (classification) built on top of Scikit-Learn
- DEA Software - pyDEA is a software package developed in Python for conducting data envelopment analysis (DEA).
- Covering Set - Constraint programming analysis.
- Insurance - CP Insurance analysis.
- Machine Learning + CP - Machine Learning + Optimisation.
- Post Office - Post Office optimisation.
- Soda - CP - Constraint Programming + ML.
- Soda - Knapsack - Knapsack algorithm + ML.
- Soda - MLP - MLP analysis + ML.

Casual Inference

- Marketing AB Testing - A/B Testing Experiment.
- Legal Studies - Instrumental and discontinuity causal approach.
- A-B Test Result - Initial A-B Results.
- Causal Regression - Regression technique for causal estimate.
- Frequentist vs Bayesian A-B Test - Comparison between frequentist and Bayesian A-B testing.
- A-B Test Power Analysis - Sample size estimation to match testing power.

- Variance Reduction A-B test - Techniques to reduce variance in A-B tests.

Statistics

- Various - Various applies statistical solutions

Quantitative

- Applied RL - Reinforcement Learning and Decision-Making tutorials explained at an intuitive level and with Jupyter Notebooks
- Process Mining - Leveraging A-priori Knowledge in Predictive Business Process Monitoring
- TS Forecasting - Time series forecasting for important business applications.

Data

- Web Scraping - Web scraping solutions for Facebook, Glassdoor, Instagram, Morningstar, Similarweb, Yelp, Spyfu, LinkedIn, Angellist.

Operations

Failure and Anomalies

- Anomalies - Anomaly detection resources.
- Intrusion Detection - Detecting network intrusions.
- APS Failure, Data - Investigating APS failures in Scania trucks.
- Hardware Failure - Using different machine learning techniques in detecting anomalies.
- Anomaly KIs, Paper - Anomaly detection algorithm for seasonal KPIs.

Load and Capacity Management

- House Load Energy - Linear, SVR and Random Forest models to predict house's appliances energy Load.
- Uber Load Management - Uber predictive load management.
- Capacity Management - Investigating IT stability issues are caused by capacity constraints.
- Bike Sharing - XGBRegressor, RandomForestRegressor, GradientBoostingRegressor combined with feature selection.
- Airline Fleet Segmentation - Analysis of Delta airlines.
- Airbnb - Airbnb Booking Analysis.

Prediction Management

- Dispute Prediction - Financial service complaint management.
- Flight Delay Prediction - Transfer learning for flight-delay prediction via variational autoencoders in Keras.
- Electric Fault Prediction - Predict tripping at grid stations by applying simple machine learning algorithms.
- Popularity Prediction in R - Marked Hawkes Point Process.