Kari Palmier DePaul University Coursework Completed

- CSC 529 – Advanced Machine Learning
    - Python numpy, pandas, matplotlib, scikit-learn, imbalance-learn used
    - Decision Tree, Random Forest, AdaBoost, Gradient Boost, Neural Network classification and regression models
    - SVM, LDA, Gaussian Naïve Bayes, Bayesian Network, and Markov temporal models
    - Chi-squared, low variance filter, mutual information, and stepwise recursive feature selection
    - PCA dimensionality reduction
    - Model performance analysis
- CSC 578 – Neural Networks and Deep Learning
    - Python Keras and Tensorflow used
    - Gradient descent of error surface
    - Artificial neural networks and back propagation
    - Activation functions (sigmoid, tanh, ReLU, Softmax)
    - Cost functions (quadratic, cross entropy, log likelihood)
    - Regularization (L1, L2, dropout layers)
    - Hyperparameters (learning rate, momentum, number of epochs, batch size)
    - Convolutional neural networks
    - Recurrent neural networks (LSTM, GRU)
- CSC 495 – Social Network Analysis
    - R and Gephi used
    - Network generation (R graph objects, adjacency matrices, edge lists, bipartite) and decomposition
    - Network metrics (edge weights, node degree, degree distribution, diameter, density, closeness, betweenness, modularity, local and global clustering)
    - Centrality measures (degree, closeness, betweenness, eigenvector, and PageRank)
    - QAP regression and exponential random graph modeling to test hypotheses about graph formation.
- CSC 555 - Mining Big Data
    - Hadoop Apache Hive, Apache Pig, Hadoop Streaming, Apache Mahout, Apache HBase, Apache Spark, and Python used
    - Creation of multi-node cluster using Amazon AWS instances
    - Installation of Hadoop HDFS, Hive, Pig, Mahout, HBase, and Spark on multi-node cluster
    - Table creation, population, and querying (including transformations) in Hive and Pig
    - Performing MapReduce operations such as queries and joins using Hadoop Streaming with mapper and reducer code written in Python
    - Table creation and population in HBase
    - Kmeans clustering using synthetic control in Mahout
    - Matrix factorization and prediction using test and training recommender data in Mahout
    - File loading, mapping, and reducing in Spark using PySpark
- CSC 425 – Time Series Analysis
    - R used
    - Normality (skewness, kurtosis, Jacque-Bera), autocorrelation and independence (Ljung Box), and stationarity (augmented Dickey-Fuller) testing
    - Autoregressive, moving average, ARMA, and ARIMA modeling and prediction
    - Data transformation using sampling, differencing and/or return calculations to provide stationarity
    - GARCH volatility modeling
- CSC 478 - Programming Machine Learning Techniques
    - Python numpy, pandas, sklearn, and matplotlib libraries used
    - Data preprocessing (missing value replacement, normalization, test/train splitting)
    - Exploratory analysis including plotting and correlation

- - Decision tree, k nearest neighbors, naïve bayes, and LDA classification
  - Linear, Ridge, Lasso, and Stochastic Gradient Decent regression
  - K means clustering and PCA
  - Apriori association rule discovery
  - Collaborative and content based filtering recommender systems
  - Matrix factorization and SVD
  - Text classification and cluster modelling
  - Random forest and Adaboost ensemble classification
- CSC 465 - Data Visualization
  - R and Tableau used
  - Line plots, scatter plots, bar charts, treemaps, hierarchical bar charts, stacked bar and area plots, pie charts, polar plots, horizon plots, level plots, boxplots, heatmaps, map chloropleths, cartograms, mosaic plots, network graphs, forced direct graphs
- CSC 424 - Advanced Data Analysis
  - SPSS used
  - Multivariate regression, PCA, factor analysis, LDA, K means and hierarchical clustering, canonical correlation
- IS 467 - Fundamentals of Data Science
  - SPSS used
  - Data cleaning, normalization, binning, sampling, aggregation, filtering, cross-validation, test/train split evaluation, decision tree and k nearest neighbor classification, k means and hierarchical clustering
- CSC 455 - Database Processing for Large-Scale Analytics
  - Oracle SQL and SQLite through Python sqlite3 API used
  - Relational database structure, table creation, table modification, constraints, queries, joins, views, schema normal form decompositions
- CSC 423 - Data Analysis and Regression
  - R used
  - Exploratory analysis through scatterplots, histograms, bar charts, and boxplots, correlation analysis, multiple linear regression, logistic regression feature selection
  - Regression analysis includes R squared and adjusted R squared, F test, coefficient significance, residual analysis, outlier, multicollinearity, and influential point analysis
- CSC 412 - Tool and Techniques for Computational Analytics
  - Matlab used
  - Linear algebra covering reduced row echelon form, linear systems, LU decomposition, orthogonalization, eigen vectors and values, SVD, least squares regression
- IT 403 - Statistics and Data Analysis
  - SPSS used
  - Descriptive statistics, boxplots, histograms, normal plots, normal distribution, z scores, quartile plots, correlation analysis, linear regression, probability, hypothesis testing