USC Marshall School of Business

DSO 528: BLENDED DATA BUSINESS ANALYTICS FOR EFFICIENT DECISIONS
Spring 2024, 16273R, 16277R
Location: JKP204,   M, W 3:30 P.M. – 4:50 P.M.
This is Syllabus Outline, the actual Syllabus will be posted before the Semester

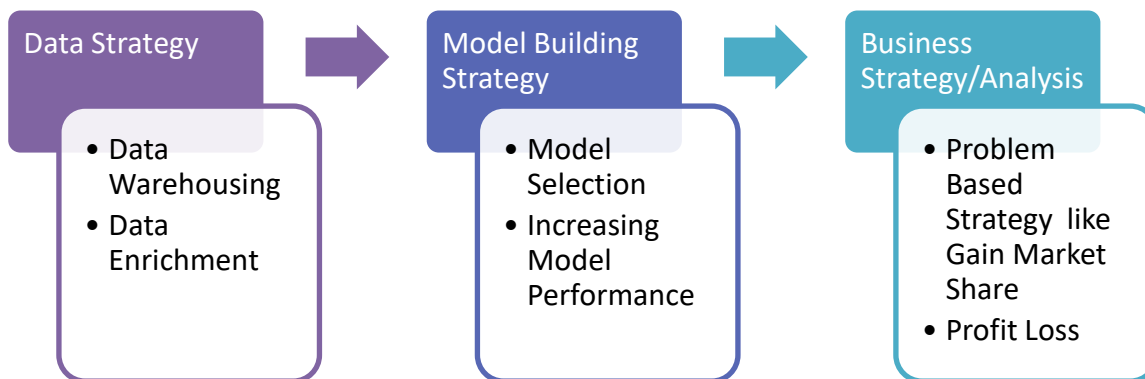|	
Instructor:	Dr. Arif Ansari
Office:	BRI 401 R (Main Campus) and Zoom Office hours
Office Hours:	5:00-5:45 p.m. on Class days and by Zoom Appointment
Phone:	(213) 999 3554
Email:	aansari@marshall.usc.edu

Emergency Contact number: 213-740-0172 or 213-999-3554

WHY TAKE THIS COURSE?
All Students who want to get ahead in the new AI and data rich world and those who plan to have a career in Business Analytics, Digital Transformation, Digital Marketing, Financial Modeling, Supply Chain Analytics, Accounting, Analytical Strategy Development and Entrepreneurs who want to build Business Models using Analytics, etc.

Today Analytics is in every field, knowing how interpret the Data, Dash Board and Model Report is critical for every employee. Learning how to solve business problems using Business Analytics is key to successful career. Learn how to build strategy driven models with Enterprise data & third party data for efficient decisions.

This course covers the entire Spectrum of developing Implementable, Interpretable, Optimized Analytic Models using Internal, External and Enriched Data that fulfills Business Strategy and Business Criteria. Understanding End-to-End Analytical Process helps you to do well in your career.

Data Strategy	Model Building Strategy	Business Strategy/Analysis
• Data Warehousing • Data Enrichment	• Model Selection • Increasing Model Performance	• Problem Based Strategy like Gain Market Share • Profit Loss

COURSE DESCRIPTION

Build Analytical Models for Classification, Clustering and Association Problems. Leverage third party "Big Data" for enriching and monetizing data. Develop data mining and business analysis skills.

LEARNING OBJECTIVES

Upon successful completion of this course, students will be able to:

1. Describe Data Warehousing (DW), Big Intelligence (BI) and Data Mining (DM) from business point of view.
2. Apply Descriptive, Diagnostic, Predictive and Prescriptive Analytics for Business problems.
3. Build Analytical Models (Data Mining) for Classification, Clustering and Association Problems using SAS JMP software.
4. Select and use the appropriate analytical techniques for analyzing the data - devising creative, sustainable, and achievable strategies and solutions that allow organizations to take advantage of opportunities that create value for its stakeholders.
5. Develop quantitative material to support written arguments - Applying critical and analytical thinking to identify viable solutions that can create short-term and long-term value for organizations.
6. Apply data mining skills to monetize data - applying an integrated approach to understanding and analyzing significant business problems, which can be complex, messy, unstructured, and beyond formulaic analysis.

COURSE MATERIALS

The following items will be necessary for completion of cases, team project and successful completion of the course.

1. Class notes and Class Videos
Class notes for this class will be available on blackboard. You should familiarize yourself with these notes before they are covered in class. In addition, I will post class videos for your preparation and I expect you to watch them before Class.

If you have any questions or need assistance with the Blackboard Course Pages, please contact the Marshall HelpDesk at 213-740-3000 or HelpDesk@marshall.usc.edu

2. Text Book is optional.

EXPECTED LEARNING OUTCOMES

- You will get to know the Big Data Analytics Domain
- You will be able to frame and solve Business Analytical problems
- You will be capable of building predictive models
- You will get hands on data mining skills to monetize data
- **6 in-depth real-world cases on Online Marketing, Search Engine Marketing, Loan Default, Churn, Ensemble Modeling, Data Mart and short cases on AI & Association**
- You will become familiar with JMP software
- You will learn to interpret and communicate the key business insights from model building
- You will become familiar with Data Warehousing
- You be able to Envision, Manage and Lead Analytical Projects (Entry Level)

COURSE STRUCTURE

- ➢ **70% of the class will be focused on Data Mining**
- ➢ **10% on Business Intelligence tools**
- ➢ **20% on Data warehousing**

Overview:

Big Data has made available a lot of Structured and Unstructured data along with it new Paradigms for Monetizing data. Many of these new paradigms are ill-defined problems and businesses are struggling to develop methods and models to leverage the "Big Data Opportunity." Most entry level Analytics professionals are "man-in-the-middle" between data and business requirements; they don't know how to fully utilize all the available data and are not fully aware of the domain expertise needed to build

efficient business savvy models. This course is for students who want to be "Standing on the Shoulders of Giants (Big Data Analytics)" and have great vision on the data side and on the business side, understand Big Data - its potential and drawbacks, Statistics - its usefulness and limitations, data mining - its usefulness and limitations, Business needs and available opportunity. In short this class is about how to monetize Data in customer facing applications using critical thinking and creative thinking.

Companies have huge amount of data in their data warehouse and have access to Big Data through 3rd party APIs. Companies want to leverage data for decision making by building "Data Driven Decision Making Models" and they want to monetize big data using data mining (DM) and Business Intelligence.

The access to social, demographic, transactional, click-stream, web usage etc., data has made companies "data rich" and now they want new ways to monetize data as wells as enhance the traditional predictive models using Enriched data. For example, Fortune 500 companies such as American Express, Wells Fargo and Wal-mart have accumulated a great deal of data from their day to day business now they want to monetize the data by providing value to customers and sell their products and services through Omni channels in an efficient manner.

In the Big Data Analytics space what are critical are information, knowledge, insight and monetization. Some of the questions are: what is the utility of the data? How can one use data in managing customer relationship and empowering employees? How can one uncover patterns and relationships hidden in databases? How can one creatively find ways to monetize data through analytical models? How can one enhance the performance of existing models?

In summary, managers need to understand the strategic values of their company's information assets, be capable of building analytical models to monetize data, understand the models built by third party companies, be able to extract insights from the models and be able to visualize data and insights.

Structure of lectures:

DSO 528 will be organized in a way that includes some combination of the following: lectures, case-based class discussion and computer work.

This class is designed in such a way that only limited mathematical and statistical (Descriptive Statistics, Hypothesis testing and Regression) background is required. I will give a brief review on the above-mentioned topics. Learning and understanding underlying DW/BD concepts, studying cases, applying DM/BD ideas and methods to business data, and communicating ideas and solutions will be our main theme. Technical details of selected DM methods will be discussed. Students are expected to use Data Mining software for various cases in class.

SOFTWARE – JMP – Refer to Blackboard for Instructions to download JMP 17
(Note: On some topics I will show how to Model using R Studio – Please note JMP17 provides an extensive output to make Business Decisions compared to Standard R without specialized packages)
Go to USC ITS website (you can also google USC JMP to get to the website) also instructions to download are provided in Blackboard.

You will be using Excel and JMP software to describe and analyze data. Excel is not a good tool for Regression Analysis, so JMP will be used for Regression Analysis. You will find the JMP software manual quite useful. JMP resembles a spreadsheet in some ways but has many specialized graphical features not found in Excel, working with JMP15 will help you to do Visual Analytics for other courses.

The JMP Manual will be used as a reference for using the JMP software and for its descriptions and discussions of statistical concepts. Note: There is both a Windows and a Mac version of the software.

If you have any questions or need assistance with the Blackboard Course Pages, please contact the

COURSE FORMAT

This course is taught in a partial flipped format. All course materials can be found in Blackboard (https://blackboard.usc.edu).This course is divided into modules, and each module is intended to cover one or two or three class session(s). Module activities may include reading assignments, weekly lessons, interactive exercises, homework assignments, Blackboard discussion forums, class sessions, and quizzes or exams. It is expected that students will have completed all required activities and assignments *before* attending their class session for each week**.**

Students should ensure that they can access all of the online tools via Blackboard prior to the start of classes.

GRADING

Your final course grade will represent how you performed in the class relative to other students. Your grade will not be based on a mandated target, but on your performance. Historically, the average grade for a graduate elective class at USC Marshall is about a 3.5 average. Your grade will be based on the following components (refer below)

Assignments	% of Overall Grade
Individual Cases & Class Participation (including written submission)	10%
Group Cases (Best 5 out of 6)	30%
Mid-Term Exam	30%
Final Exam	30%
TOTAL	100%

CLASS PARTICIPATION
Class participation is an extremely important part of the learning experience in this course as the richness of the learning experience will be largely dependent upon the degree of preparation by *all* students prior to each class session. In addition to class discussion, you will submit your answer in written format via Blackboard Individual Discussion. Refer to Appendix III

Group Cases
Professor Ansari encourages to form groups based on the students' interest, making sure the groups are made **up of students from different degree programs**. We will analyze six cases during the semester. The cases will be evaluated and will be counted towards the case points. The Cases can be done in groups of 3 or 4 students. **Your Team will submit one case report for each case. Your team will do a peer evaluation for each case and the peer evaluation is part of the grade.**

EXAMS
Exams are open book. You need to a business calculator that can do business calculations. Suggested calculator is HP 17B.

The Midterm will cover material taught in the first half of the class. You CANNOT be exempted from the Midterm. NO MAKE-UPS OF Midterm WILL BE GIVEN.

The Final Exam will cover only material taught in the second half of the class. You CANNOT be exempted from Final Exam even if you have an A average. NO MAKE-UPS OF Final Exam WILL BE GIVEN.

I. The course will start with Data Mining. The Data Mining part of the class will be quantitative and the following topics will be covered in it.

      1.   Standard Data Mining techniques:
      a.   Classification
      b.   Clustering
      c.   Association
      d.   Visual Data mining

Using various appropriate techniques,
i)   Decision Tree
ii)   Logistic Regression
     iii)   Neural Networks
     iv)   Naïve Bayesian
     v)   Nearest Neighbor
     vi)   Other techniques like Ensemble Modeling
vii) Partitional and Hierarchical Clustering

II. The second part of the course will be Data Blending and Data Warehousing. You will be introduced to Data Warehousing from business perspective, how to create Data Warehouse Architecture and other sources of data for Data Blending.

Information on Cases

Case 1 – Prof. Ansari's Smart Partyware – "Find the high propensity customers for Celebrating American Arts product"

Learn Key Concepts like Perfect Marketing, Decision Tree, Beating Computer Models, Understanding Data Mining Metrics, KPIs, Lift and Monetization

The Smart Partyware Company's business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

In the direct-marketing industry, the response rate is measured as a percentage of customers who buy the directly mailed product. Smart Partyware's historical response rate for direct mail to selected members is approximately 10%—far above the industry average. SPW was using RFM (Recency-Frequency-Monetary) analysis to target customers. Smart Partyware wants to increase the response rate well beyond the 10% rate.

SPW designs new party ware for every campaign, gives a new name to its party ware, and broadly classifies the party ware under one of its many party themes. Most of the designs cut across many themes but are classified into a particular category based on the main design theme in the party ware. The recent product to be marketed is Celebrating American Arts. It has famous American art works printed in the party ware and the objective is to find high propensity customers for the current marketing campaign. In this case the students will be using Decision Tree Model.

Case 2 – Prof. Ansari's Search Engine Marketing for Smart Partyware – "Selecting the right set of keywords for search engine marketing campaign.

Learn Key Concepts like Search Engine Marketing, Dash Boards, Clustering and Campaign Management

The Smart Partyware (SPW) Company's business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

Applichem has signed a Memorandum of Understanding (MOU) with SPW. They will acquire 10 percent of SPW for an undisclosed sum and have an option to buy up to a total of 49% in the following year at current valuation determined by independent evaluators.

John Runner one of the founders of SPW has a vested interest¬—he wanted to increase the revenue and profit of SPW so that the valuation of SPW in a year will be high and Applichem will have to pay more for the shares of SPW. John and other executives' contracts with SPW allowed them to sell up to 25% of their shares as part of the deal with Applichem. John Runner was sure his prodigy Vijay would be able to do his magic once again and would be able to increase revenue and profit.

Vijay knew he had fully leveraged the power of data mining; increasing the efficiency of the algorithms would not increase the revenue and profit by 50%. His first approach was to buy a potential member list from data brokers to increase the number of members at SPW. This approach was not successful; the additional revenue from new members was not substantial. In fact, the profit from new members was negligible after taking into account the amount of money paid for the data acquisition and the cost of phone-based marketing to enroll them as new members. The second approach was revamping the site and doing Search Engine Marketing (SEM).

SPW signed up with Google AdSense and created an account with Google. Based on "Partyware" search wording Google AdSense gave a list of nearly 800 keywords and phrases that people normally search, along with the level of competition, the number of local monthly searches, and the approximate cost per click (CPC). The total amount spent per month on the "partyware" keyword was approximately $250,000. SPW agreed to allocate $20,000 for ad budget the first month, and based on the success or failure the next month's budget would be decided. Based on the keyword bidding SPW wants to sign-up as many visitors to its website as members and increase its membership base.

The challenge of this case is to find the right cluster(s) of keywords for SEM campaign so that many prospects will visit the website and join as members.

Case 3 – SAS's "Home Equity"

Learn Key Concepts like Logistic Regression, Profiler, Odds, KPIs, Lift and Monetization

The consumer credit department of a bank wants to automate the decision-making process for approval of home equity lines of credit. To do this, they will follow the recommendations of the Equal Credit Opportunity Act to create an empirically derived and statistically sound credit scoring model. The model will be based on data collected from applicants granted credit through the process of loan underwriting. The model will be built from predictive modeling tools, but the created model must be sufficiently interpretable so as to provide a reason for any adverse actions (rejections).
The HMEQ data set contains baseline and loan performance information for 1000 recent home equity loans. The target (BAD) is a binary variable that indicates if an applicant eventually defaulted or was seriously delinquent. This adverse outcome occurred in approximately 10% of the cases. The challenge of this case is to predict the BAD loans so the consumer credit department will be able to prevent default and make appropriate decision on the home equity line of credit.

Case 4 – New Case Churn using Neural Network

Learn Key Concepts like Neural Network, Transformation, Profiler and Leveraging Multiple DM methods

Customer retention is a challenge in the ultracompetitive mobile phone industry. A mobile phone (service provider) company is studying factors related to customer churn, a term used for customers who have moved to another service provider.

The Task
The company would like to build a model to predict which customers are most likely to move their service

Currently there are 1 million customer accounts, this month the expected churn rate is 15%. The monthly revenue is $50 per customer per month. It will cost $10 discount per month to keep customer who might churn away from churning. The Current Revenue is $50 million. If all the people who want to churn leave the company then the revenue will be $42.751 Million.

Assume the company is willing to review up to 15% accounts that they think will churn and give $10 proactive discount of $10 (promotion).

It is important to correctly identify the potential churner among the million customers.

Data-Mining Business Models

The mobile company believes the best method to use is Neural Network as they will be able to defend the model based on non-linear modelling.

Help Build a model to identify the churners and maximize the new net revenue per month (refer to excel sheet for calculations).

Case 5 – Prof. Ansari's Smart Partyware – Big Data Approach – Data Blending and Ensemble Modeling

Learn Key Concepts like Big Data, Data Blending, Ensemble Modeling, SAS-Enterprise Miner Tools

The Smart Partyware Company's business model is direct-to-consumer marketing. Over the years they have gained dedicated upscale customers and currently have 500,000 members in their database.

In the direct-marketing industry, the response rate is measured as a percentage of customers who buy the directly mailed product. Smart Partyware's historical response rate for direct mail to selected members is approximately 10%—far above the industry average. SPW was using RFM (Recency-Frequency-Monetary) analysis to target customers. Smart Partyware wants to increase the response rate well beyond the 10% rate.

SPW designs new party ware for every campaign, gives a new name to its party ware, and broadly classifies the party ware under one of its many party themes. Most of the designs cut across many themes but are classified into a particular category based on the main design theme in the party ware. The recent product to be marketed is Celebrating American Arts. It has famous American art works printed in the party ware and the objective is to find high propensity customers for the current marketing campaign. In this case the students will be using the standard business analytics methods like, Decision Tree, Logistics Regression

and Neural Network.

Case 6 – Star Schema – Model Trojan Dental School Business Intelligence Engine.

Learn Key Concepts like Data Mart, Star Schema, KPIs and BI
Trojan Dental School is involved in continuing education program for Dentists. This is revenue generating area for the Dental School and they want to increase the revenue and efficiency in targeting the potential candidates for continuing education.

The Dental School has Alumni data and can buy the list of Dentists in Southern California area from a data-broker. You are hired by Trojan Dental School to create Business Intelligence Engine that will enable the dental school to become the premier continuing education program for Dentists in Southern California region, to increase revenue from its Continuing Education Program, to estimate its Customer Life time value, to improve its Customer Service, to Create actionable dashboards for its Administrators and faculty and make their continuing education program an agile organization.

Inclass Cases:

Lost Sales: Find out the reasons why a Business Lost Bidding using different Analytical Model,

Movie Recommendation: Join different Dataset on Movie and build Association Algorithm for Movie Recommendation

Demo for Learning Purpose:

Reinforcement Learning: How to build Agents using Matlab

Deep Learning using R

DSO 528 – **Blended Data Business Analytics for Efficient Decisions**

Section(s) – 16273/16277

Professor

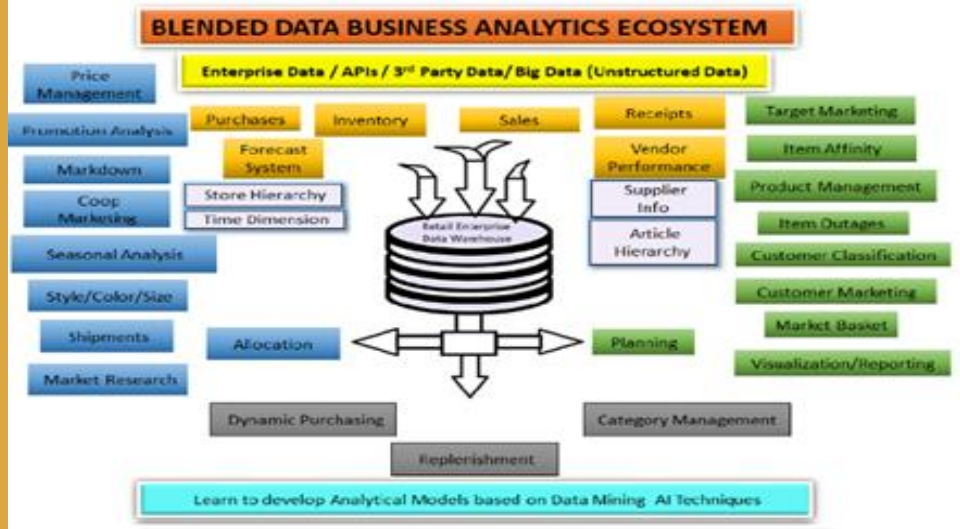*Arif Ansari*

Email

aansari@marshall.usc.edu

When

Mon/Wed, 3:30 OM – 4:50 PM

Office	**Units**
BRI 401 R	*3.0*

WHY TAKE THIS COURSE?

All Students who want to get ahead in the new AI and data rich world and those who plan to have a career in Business Analytics should take this class. Students. Today Analytics is in every field, knowing how interpret the Data, Dash Board and Model Report is critical for every employee. Learning how to solve business problems using Business Analytics is key to successful career. Learn how to build strategy driven models with Enterprise data & third party data for efficient decisions.

COURSE OBJECTIVES

- To provide students with concepts, frameworks, analytical thinking, critical thinking and creative thinking skills for converting Enterprise Data and Big Data into analytical models for monetizing data.
- To provide practical knowledge (six in-depth cases), skills, methods, tools, KPIs and resources for conceiving, building and solving business problems using Analytical Models.
- Provide End-to-End approach to solving Business Problems, Data Strategy → Analytics → Business Analytics → Business Analysis

KEY CONCEPTS

- Data Mining/Business Intelligence/AI
- Blended data/Data Warehousing/Enrichment
- JMP Software for rapid model building
- JMP/R-studio comparison for better reporting
- Descriptive/Diagnostic/Predictive/Prescriptive /Investigative Analytics
- Prediction, Classification, Clustering & Association
- Decision Tree, Logistic Regression, KNN, Neural Network and Ensemble Model.
- Introduction to CNN/RNN/Generative AI/Random forest/ Naïve Bayesian/SVM
- Partitional and Hierarchical Clustering
- Search Engine Marketing (SEM)
- Star Schema, Dash Boards, Optimizer.

COURSE DESCRIPTION

The course focus is to give a Big Picture view of Business Analytics, its components and platforms. To build sophisticated strategy driven business analytical models from raw data using Desk top and Industry level tools for Classification, Clustering and Association Problems. To show how to leverage the readily available "Big Data" from third party sources for enriching and monetizing data. To develop data mining and business analysis skillset to gain inference from your analysis, from Executive, Business and Statistical point of view. To provide a systematic approach to build Analytical Models. To provide the missing link between Analytics and Business Analysis.