University of North Florida - School of Computing
Course Syllabus for

CAP 6768 – Data Analytics

(Graduate – 3 Semester Credits – Online)

Instructor

Karthikeyan Umapathy, Associate Professor
Office location: Mathews Building 15, Room 3214
Office hours: by appointment only
Phone: 904-620-1329 (office)
Email: k.umapathy@unf.edu (preferred communication method)

Important Dates

Team Information: **Week 2**
Group Contract: **Week 3**
Survey Paper – Topic Selection: **Week 4**
Course Withdrawal (25% Refund): **Week 4**
Data Mining Project Deliverable 1 – Proposal: **Week 6**
Survey Paper – First Draft: **Week 8**
Data Mining Project Deliverable 2 – Data Collection Report: **Week 10**
Data Mining Project Deliverable 3 – Data Preparation Report: **Week 12**
Course Withdrawal (No Refund): **Week 12**
Survey Paper – Interim Draft: **Week 13**
Symposium Poster File: **Week 15**
SOC Symposium – Data Mining Project Presentation: **Week 15**
Data Mining Project Final Report: **Week 16**
Survey Paper – Final Draft: **Week 16**

Syllabus Contents

Course Information

Catalog Description

The aggressive rate of data growth has outpaced our ability to manually understand what data represents. Data is typically stored in database and files, and represented in different formats (structured, semi-structured, or no structure). Data analytics is the science of applying quantitative techniques to analyze data with the objective of discovering hidden knowledge and identifying interesting patterns. This course surveys a number of data preprocessing and sampling methods, data distributions and uncertainty, statistics, regression, time-series analysis, predictions and clustering. It introduces the characteristics and analytic challenges on dealing with clinical data from electronic health records. The course also covers emerging trends in Data Analytics and the

applications of information technology in the healthcare. Statistical analyses and data mining techniques will be discussed along with methods for deploying these techniques using the open source tools.

Textbooks

Course slide materials, quizzes, and some of the discussion activities will be based on the selected chapters from below two textbooks.

1. Business Analytics: Data Analysis & Decision Making by S. Christian Albright and Wayne L. Winston. 6th Edition. Published by Cengage Learning. (ISBN: 9781305947542)

2. Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, and Jian Pei. 3rd Edition. Published by Morgan Kaufmann. (ISBN: 9780123814791)

Instructor's Perspective About the Course

We are living in the age of technology that has enabled the ability to collect data on virtually every aspect of business as well as our personal and social activities. The always-connected world has produced a vast amount of data and will continue to generate an enormous amount of data. The board availability of data has led to the challenge of developing efficient methods for extracting useful information from the data and effective processes to convert that information into actionable knowledge. Organizations in almost every industry are conducting quantitative analysis on data to make data-driven decisions. Organizations can gain competitive advantage by employing quantitative methods to uncover information in data sets and then act on information to obtain an advantage that their competitors are not able to gain. The data science discipline was formed to take advantage of growing data and address the challenge of extracting actionable insights from collected data.

One of the fundamental topics in data science is data mining techniques which are used for analyzing information and extracting knowledge contained within datasets. Data mining techniques help us analyze heterogeneous, noisy, unstructured, and large-scale datasets to discover hidden patterns that could provide valuable insights. In this is a graduate-level course, you will gain an overview of the data analytic process and data mining techniques used for discovery knowledge from datasets. The course is designed to highlight the practical aspects of data mining methods and their applications, rather than theoretical aspects of statistical machine learning or optimization. The course materials will focus on how the information in different real-world problems can be formulated, and how the basic mining tasks can be performed. At the end of the course, you should walk away with knowledge of the basic concepts and methods of data mining as well as how a data analytic problem should be approached, i.e., how to process and prepare the data for analysis, and which mining technique should be adopted to solve a given problem.

Learning Outcomes

Upon completion of the course, students should be able to:

- Define the term "Analytics" for healthcare.
- Discuss the steps involved in the data mining process (e.g., preprocessing, classification, regression, clustering, and visualization) and apply them for analysis of healthcare data.
- Describe different methods of predictive analytics and their applications in the healthcare domain.
- Apply data preprocessing techniques for healthcare data.
- Be proficient with analytics tools for healthcare data preparation and analysis.

- Apply data mining techniques to healthcare problems.
- Evaluate the data from diverse sources to create meaningful presentations.
- Create business intelligence for healthcare through data analytics.

Method of Teaching

This course is a highly interactive class that incorporates online teaching methods. This course utilizes a variety of teaching mediums including the textbooks, video lectures, PowerPoint notes, student public presentations, team project work, team project reports, individual project work, individual writing assignments, discussion forums, and quizzes. As with all university courses, expect to dedicate at a minimum of 9 hours of study per week into this course. The nature of online education requires students to be actively involved with and take more responsibility for their learning. This distance learning course is asynchronous (students may log on to Canvas at any time to complete coursework). However, students must participate in course discussions and submit their work on time by the due date outlined in the course schedule.

Student Responsibilities

Everyone, especially those new to online courses or Canvas, should review UNF's resources and tips on taking Distance Learning (DL) classes at http://www.unf.edu/distancelearning/current/FAQ.aspx. If you have technical problems or a question on how to use Canvas, call the Help Desk at 620-HELP or check web-link above. If you cannot resolve the issue, go to a campus computer helpdesk (available at Library). It is your responsibility to understand how to use Canvas correctly.

Configure your personal computer to navigate correctly and efficiently on Canvas by the end of the first week of class or plan to use a UNF computer. Use a current, standards-compliant web browser such as Google Chrome or Mozilla Firefox.

Method of Evaluation

Assessment Item	Team Assessment	Individual Assessment
Data Mining Project		
Deliverable 1 – Proposal	1%	4%
Deliverable 2 – Data Collection Report	5%	10%
Deliverable 3 – Data Preparation Report	5%	10%
Deliverable 4 – Final Report	5%	10%
Project Presentation at Symposium	5%	5%
Survey Paper		
First Draft		5%
Interim Draft		5%
Final Draft		10%
Class Participation		
Quizzes		10%
Discussion Forums		10%
Sub Total	21%	79%
Total	100%	

Letter grades will be based on:
94 – 100 = A
90 – 93.99 = A-
87 – 89.99 = B+
84 – 86.99 = B
80 – 83.99 = B-
77 – 79.99 = C+
70 – 76.99 = C
60 – 69.99 = D
less than 60=F

The penalty for cheating or plagiarizing on assignments will be F grade in the course. Work which is similar beyond coincidence will automatically be considered cheating by all parties.

Late Assignments
There will be a penalty of 10 % per day for late submission of assignments (including weekends and holidays).

Course Passing Requirement

In order to pass this course, at minimum 50% of grade points must be obtained for each assessment item. If you received less than 50% grade points for one of the assessment item, then you will receive F as the final letter grade.

Academic Dishonesty

UNF will not tolerate academic dishonesty in any form as it is contrary to the process of learning. Students should demonstrate academic integrity in all of their course works. Students who violate university rules on academic dishonesty will be punished with the most severe penalty allowed by the university policy.

Please review the University policy on academic misconduct at:
https://www.unf.edu/president/policies_regulations/02-AcademicAffairs/EnrollmentServices/2_0640P.aspx

The policy on academic integrity and misuse of computer equipment and computer accounts found at http://www.unf.edu/ccec/computing/Policies___Guidelines.aspx.

Violations of Academic Integrity
Under this heading the University of North Florida Student Handbook identifies several types of violations; these include but are not limited to: cheating; fabricating and falsifying information or citations; submitting the same work for credit in more than one course; plagiarizing; providing another student with access to one's own work to submit under this person's name or signature; destroying, stealing, or making inaccessible library or other academic resource material; and helping or attempting to help another person commit an act of academic dishonesty. The University of North Florida authorizes any instructor who finds evidence of cheating, plagiarism, or other wrongful behavior that violates the University of North Florida Academic Integrity Code to take appropriate action. Possible action includes, but is not limited to, failing the student on the work in question, failing the student for the course, notifying the appropriate academic dean or Vice President for Student Affairs,

and requesting additional action be taken. The consequences of a breach of academic integrity may result in an F, which is unforgivable, regardless of withdrawal status.

Matching Assessment Items to Learning Outcomes

Course outcomes	Assessment Items
Define the term "Analytics" for healthcare.	Quizzes, Discussion Forums, Survey Paper First Draft, Survey Paper Interim Draft, and Survey Paper Final Draft
Discuss the steps involved in the data mining process (e.g., preprocessing, classification, regression, clustering, and visualization) and apply them for analysis of healthcare data.	Discussion Forums, Data Mining Project Deliverable 4 – Final Report, and Data Mining Project Presentation
Describe different methods of predictive analytics and their applications in the healthcare domain.	Survey Paper First Draft, Survey Paper Interim Draft, and Survey Paper Final Draft
Apply data preprocessing techniques for healthcare data.	Data Mining Project Deliverable 2 – Data Collection Report, and Data Mining Project Deliverable 3 – Data Preparation Report
Be proficient with analytics tools for healthcare data preparation and analysis.	Data Mining Project Deliverable 1 – Proposal, Data Mining Project Deliverable 2 – Data Collection Report, Data Mining Project Deliverable 3 – Data Preparation Report, Data Mining Project Deliverable 4 – Final Report, and Data Mining Project Presentation
Apply data mining techniques to healthcare problems.	Data Mining Project Deliverable 4 – Final Report
Evaluate the data from varying sources to create meaningful presentations.	Data Mining Project Deliverable 4 – Final Report, Data Mining Project Presentation, Survey Paper First Draft, Survey Paper Interim Draft, and Survey Paper Final Draft
Create business intelligence for healthcare through data analytics.	Data Mining Project Deliverable 1 – Proposal, Data Mining Project Deliverable 2 – Data Collection Report, Data Mining Project Deliverable 3 – Data Preparation Report, Data Mining Project Deliverable 4 – Final Report, and Data Mining Project Presentation

Deliverables

Data Mining Project

Data mining is a cornerstone of analytics that helps you develop the models that can uncover connections within millions or billions of records. Data mining allows you to sift through all the chaotic and repetitive noise in your data and understand what is relevant and then make good use of that information to assess likely outcomes. Data mining represents a variety of (descriptive, predictive, and prescriptive) models used in different analytic capabilities that address a gamut of organizational needs, ask different types of questions and use varying levels of human input or rules to arrive at a decision. This project will involve working with healthcare datasets, solving problems using data mining techniques, presenting the analytics solutions, and writing technical reports.

Team Experience (*Team info due: Week 2*)
Students are expected to form a team of 2 or 3 students for the data mining project. Submit team name and member info.

Group Contract (*Group contract due: Week 3*)
Students are expected to create a contract that describes rules of engagement for the team.

Deliverable 1 – Proposal (*Due: Week 6*)
The project proposal briefly describes the dataset, repository from where the dataset will be obtained, define the problem the intends to solve, and data mining techniques intend to use.

Deliverable 2 – Data Collection Report (*Due: Week 10*)
The report should describe procedures followed to collect the data, hurdles faced with data collection, data format, dataset size, how the data is stored, characteristics of the data, descriptive statistics of the data, and attributes that relevant to the selected problem.

Deliverable 3 – Data Preparation Report (*Due: Week 12*)
The report should describe the process followed to explore and prepare the data for analysis using data mining techniques. In the data exploration section, describe hypotheses formed for solving the problem, attributes selected for analysis, procedures followed to verify the quality of the data, and clean the data.

Deliverable 4 – Final Report (*Due: Week 16*)
The final report should provide an executive summary, introduction, data collection report, data preparation report, data mining report, conclusion, reference, and appendix. The report should describe how the problem was solved using data analytic approaches.

School of Computing Symposium

CAP 6768 students will present their data mining project work as a poster at the School of Computing Symposium this fall will be held on Friday, December 6, 2019, from 3 PM to 6 PM. Note that you should plan to arrive at the University Center around 2 PM to have sufficient time to check-in and set up your poster.

Poster file due date: Week 15
Poster presentation date: Week 15

Please note that if you do not present at the symposium, you will not receive any project presentation points. You must be present at the symposium for the entire three-hour period from 2 PM to 6 PM, to receive presentation points. If you are not able to present the poster at the symposium event, you must reach out to the instructor to make alternate arrangements during the first week of the semester.

Data Analytics Survey Paper

Students are expected to write a survey paper that reviews the literature on a particular topic relevant to data analytics and healthcare domain. The main objective of the assignment is to summarize and critically evaluate the literature to establish current knowledge of a topic. There are no particular length requirements but suggest to aim for 3000 words. The paper must use a proper academic writing style. References should be listed in a consistent format, and citations to references should be given throughout the paper to back up facts and claims. The survey paper should consist of the following sections: abstract, introduction, background, literature review, contributions, and conclusions.

Topic Selection (*Due: Week 4*)
Submit following information in regards to selected topic for the survey paper: briefly describe specific topic that will be focus of the paper, explain why this topic is relevant to the course, briefly describe indicate to which

conference you intend to model the paper to it meets expected academic standards, and finally, explain why you selected this conference.

First Draft (*Due: Week 8*)
First draft submission must contain the following sections: title, abstract, introduction, and background.

Interim Draft (*Due: Week 13*)
Interim draft submission must contain the following sections: update the sections submitted in the first draft, and literature review.

Final Draft (*Due: Week 16*)
Final draft submission must contain all expected sections fully completed and must be updated as per remarks provided by the instructor on previous draft submissions. Expected sections are title, abstract, introduction, background, literature review, contributions, and conclusion.

Class Participation

It is essential to regularly log into Canvas, read the required materials, watch video lectures, perform project activities, perform research activities, and contribute to the class discussion forums. Participation in all of the class activities is critical for maximizing your learning in the online environment. You are expected to be an active participant in the class. Your class participation will be assessed for quantity and quality of the contributions. Class participation is assessed predominantly using quizzes and discussion forum activities.

Quizzes

Each course topic will have an associated quiz activity to test your knowledge gained. Each quiz is worth one grade point. You will be provided with two attempts to complete the activity. Final attempt score will be used towards grade calculation. Maximum obtainable grade points via quizzes (for final grade calculations) is 10.

Discussion Forum Participation

Each course topic will have an associated discussion forum to encourage discussions among students on the topic. It should be noted that participating in the self-introduction discussion forum in the first week is mandatory. If you do not participate in this forum, you will lose one grade point. Below guidelines and rubrics[1] will be used for evaluating your contributions in the discussion forums.

Guidelines for participation in the forums:

- You should submit your initial post(s) early in the session and post subsequent responses to the posts of other learners at timely intervals within the duration of the session. Keep in mind the goal is to have a dynamic discussion that lasts throughout the entire session.
- Your posts and responses should be thorough and thoughtful. Just posting an "I agree" or "Good idea" will not be considered adequate. Support your statements with examples, experiences, or references. You are, however, encouraged to be brief — keep each post and response to one or two short paragraphs. Keep in mind that your fellow learners will be reading and responding to you, too.
- Make sure to address the discussion questions posed by the instructor. This does not mean you should not extend the topic, but do not stray from the topic.

[1] http://www.pbs.org/teacherline/courses/common_documents/disc_assess.htm

- Discussions occur when there is dialogue. So, build upon the posts and responses of other learners to create discussion threads. Make sure you revisit the discussion forum and respond (if necessary) to what other learners have posted to your initial responses.
- When relevant, add to the discussion by including prior knowledge, work experiences, references, Web sites, and other related resources. (giving credit when appropriate).
- Your contributions to the discussions (posts and responses) should be complete and free of grammatical or structural errors.

Rubrics for evaluating discussion forum participation:

Unacceptable 0 points	Good 0.5 points	Excellent 1 point
Criteria: Quantity and timeliness		
• Does not create an initial post. • Does not submit early in the session. • Does not reply to another learner. • Participates not at all, or Participates 1-2 times on the same day.	• Creates an initial post. • Replies to another learner. • Participates 3-4 times but postings not distributed throughout the week.	• Creates an initial post. • Submits early in the session. • Replies to more than two learners. • Continues to participate in discussion threads until the end of the session. • Participates more than 4 times and distributed throughout the week.
Criteria: Demonstrates knowledge and understanding of content and applicability to professional practice		
• Post has little or nothing to do with the main topic or restates the central concept. • Post does not demonstrate evidence of knowledge and understanding of course materials and content. • Response does not demonstrate evidence of knowledge and understanding of course materials and content. • Does not respond to questions posed by the instructor.	• Post relates to the main topic, but no details and examples are given and have misspellings and grammatical mistakes • Post demonstrates some evidence of knowledge and understanding of course materials and content. • Response demonstrates some evidence of knowledge and understanding of course materials and content. • Responds to questions posted by instructor but does not engage in premise reflection.	• Post relates to the main topic and adds new concepts, information. It includes several supporting details and examples, and it is free of misspelling and grammatical errors • Post demonstrates clear evidence of knowledge and understanding of course materials and content. • Response demonstrates clear evidence of knowledge and understanding of course materials and content. • Response elicits responses and reflection for others. • Response to instructor questions reflects critical thinking process by

Unacceptable 0 points	Good 0.5 points	Excellent 1 point
		integrating multiple views or provides authenticated resources to take the discussion deeper.

Course Schedule

It is expected that the student will log into Canvas multiple times in a week to perform required course activities. Students should watch the video lectures and other relevant materials uploaded to the Canvas. Students should come prepared to the virtual class meetings with questions for the instructor on the course topics and project-related issues. The student is responsible for all topics presented regardless of their coverage.

Please note that below listing of chapters does not mean that all text in those chapters would be covered in this course. Only materials pertaining to class would be covered. Throughout the course, the instructor would provide other supplementary materials to provide targeted guidance to team project deliverables.

Week	Topics	Chapters	Due Dates
1	Course Introduction and Syllabus Review Introduction to Data Analytics	TB1 - Chapter 1, and TB2 – Chapter 1	Quiz 1 Discussion 1
2	Getting to Know your Data	TB2 – Chapter 2	Team Information Quiz 2 Discussion 2
3	Data Preprocessing	TB2 – Chapter 3	Group Contract Quiz 3 Discussion 3
4	Exploring Data	TB1 – Chapters 2 and 3	Survey Paper – Topic Selection Quiz 4
5			Discussion 4
6	Probability and Uncertainty	TB1 – Chapters 4 to 6	Project – Deliverable 1 Quiz 5
7			Discussion 5
8	Regression Analysis	TB1 – Chapters 10 to 12	Survey Paper – First Draft Quiz 6
9			Discussion 6
10	Mining Frequent Patterns, Associations, and Correlations	TB2 – Chapter 7	Project – Deliverable 2 Quiz 7 Discussion 7
11	Classification and Prediction	TB2 – Chapters 8 and 9	Quiz 8
12			Discussion 8 Project – Deliverable 3
13	Clustering and Outlier Analysis	TB2 – Chapters 10 to 12	Survey Paper – Interim Draft Quiz 9 Discussion 9
14	Text Mining	TB2 – Chapter 13	Quiz 10 Discussion 10
15			Project – Symposium Poster File

Week	Topics	Chapters	Due Dates
	Project Presentations		SoC Symposium
16	Final Project Report and Survey Paper		Project – Deliverable 4 Survey Paper – Final Draft

Legends
TB1 - Business Analytics: Data Analysis and Decision Making by Christian Albright and Wayne Winston
TB2 - Data Mining Concepts and Techniques by Jiawei Han and Micheline Kamber

Course Modifications

The instructor reserves the right to modify the course, including schedule, assignment specifications, assignment score distributions, grading criteria, and other relevant aspects to meet the student's needs or due to unexpected events.

Other Remarks

Students with Disabilities

Students with disabilities who seek reasonable accommodations in the classroom or other aspects of performing their coursework must first register with the UNF Disability Resource Center (DRC) located in Building 57, Room 1500. DRC staff members work with students to obtain required documentation of disability and to identify appropriate accommodations as required by applicable disability laws including the Americans with Disabilities Act (ADA). After receiving all necessary documentation, the DRC staff determines whether a student qualifies for services with the DRC and if so, the accommodations the student requires will be provided. DRC staff then prepares a letter for the student to provide faculty advising them of approved accommodations. For further information, contact the DRC by phone (904) 620-2769, email (drc@unf.edu), or visit the DRC website (http://www.unf.edu/drc/).

Military and veteran students who return from combat exposure may be utilizing the post 9/11 GI bill to continue post-secondary education goals and may need both physical and academic accommodations. Contact Ray Wikstrom, Director of Military and Veterans' Resource Center by phone (904) 620-2655, email (ray.wikstrom@unf.edu).

Student Health and Wellbeing

The UNF counseling center can help students who are having difficulties managing stress, adjusting to college, or who are feeling sad and hopeless. You can reach the UNF counseling center at http://www.unf.edu/counseling-center/ or by calling (904)-620-2602 during and after business hours for routine appointments or if you or someone you know is in crisis. Walk-in hours are 10-2 PM Monday-Friday. Crisis appointments are available.

The Victim Advocacy Program provides confidential crisis intervention services to anyone in the UNF community impacted by crime or victimization. Services offered range from emotional support to assistance navigating the criminal justice system. The Victim Advocacy Program is located in Founders Hall, Building 2, Suite 2100.

Students may contact the 24-hour Crisis Helpline at (904) 620-1010. For more information visit http://www.unf.edu/womens-center/Victim_Advocacy.aspx

Satisfactory Progress Policy

The School of Computing enforces the "one repeat" rule for all prerequisite and core courses offered by the School for its major programs. Students who do not successfully complete a prerequisite or core requirement for a School of Computing course on the first attempt (i.e., earn a grade of D, F, W, WP or WF) will be granted one chance to repeat the course. Students who do not successfully complete a prerequisite or core requirement within two attempts will not be permitted to register for courses offered by the School in future semesters. This stipulation applies whether or not the student has declared a major in a School of Computing program. http://www.unf.edu/ccec/computing/PoliciesGuidelines/Satisfactory_Progress_Policy.aspx

Continuity of Instruction Plan

In the event of disruption of normal classroom activities due to an emergency such as hurricane, pandemic or other unforeseen event or combination of events, the format of this course may be modified in order to enable completion of the course requirements. In that event, you will be provided an addendum to this syllabus that will supersede this version. It is your responsibility as a student participant to be proactive during any emergency to find instructions that I will post on Canvas which you should check daily.

CAP6768
Data Analytics