**Towson University**

**Department of Computer & Information Sciences**

**COSC880/AIT 880**

**Student’s Name:** Mary J. Snyder **Student’s No:** 0585035

**Address:** 5222 Glenthorne Ct. **Phone:** (765) – 714 – 7767

**City:** Rosedale **State:** MD **Zip:** 21237 **E-mail:** msnyde8@students.towson.edu

**Tentative Project Title:** Upset Potential in the National Football League (NFL)

**Graduate Credits Completed:** 27

**Semester to Register:** Fall X Spring\_\_\_\_ Summer\_\_\_\_ Year 2016

**Advisor’s Name:** Michael P. McGuire

**Project Description/Abstract:**

Fantasy leagues, spread picks, as well as confidence points picks have become popular among fans of a variety of sports today. While the ultimate goal of these activities is to choose all winners, the place to gain a competitive edge over opponents is in picking upsets correctly. Nowhere is this more apparent than in confidence points picks. Placing too much confidence on a game that has a surprise upset could harm the overall score more than picking a few low confidence games incorrectly.

Analysis on upset potential can be used not only in these fun situations, but the information can also help teams better prepare for games. Knowing how factors such as time zone changes, number of days rest between games, overtime games played, weather, or key player injuries can affect the outcome of the game would help the teams prepare themselves against an upset. This information can also be used to a team’s advantage by providing them ways to focus their training to upset other teams.

In this project, I will examine past NFL stats and data to determine the factors that contribute to upset games. Using the knowledge I gain from the analysis, I will attempt to predict upsets those games during an ongoing NFL season.

**Project Topics to be Studied**:

Exploratory data analysis; outlier detection; binning

Data classification

Pattern evaluation; frequent itemset mining, association, and correlation;

**Project Objectives:**

The goal of this project is to determine factors that influence the potential for upsets in the National Football League (NFL) through data mining techniques. Upon successful determination of these factors, I will attempt to predict upsets in historical data as well as predict upsets for current/real-time games.

**Deliverables:**

Project Proposal – scope and plans for graduate project

Dataset – compiled data (from many sources) used for analysis of upset potential

Source code (R file) – analysis performed on data (using R)

Written Report – Conclusion of results in the style of a conference publication

Public Presentation Slides – Summary of research and outcome, in a clear concise and presentable matter

**Anticipated Project Timeline:**

04-May: Initial project discussion

27-May: Proposal submission

29-Aug: Progress meeting 1

30-Aug: Begin research of potential attributes

12-Sep: Initial data acquisition

19-Sep: Progress meeting 2

26-Sep: Start analysis of attributes and algorithm testing/tuning

10-Oct: Progress meeting 3

27-Oct: Live/real-time testing to begin

09-Nov: Progress meeting 4

14-Nov: Complete compilation of analysis/results

02-Dec: Final report submission

Dec: Public Presentation

**Reading/References List:**

1. Weather/climate data:

National Centers for Environmental Information (NCEI), formerly known as National Climatic Data Center (NCDC). Retrieved May 20, 2016 from <http://www.ncdc.noaa.gov/>

1. Historical game results:

All Game Scores in Pro Football History | Pro-Football-Reference.com. Retrieved May 20, 2016 from <http://www.pro-football-reference.com/boxscores/game_scores.cgi>

1. Box score game results (similar results for each team/game/year):

Miami Dolphins at New England Patriots – December 7th, 2003 | Pro-Football-Reference.com Retrieved May 20, 2016 from <http://www.pro-football-reference.com/boxscores/200312070nwe.htm>

1. Historical team schedules (similar results for each team and multiple years):

2002 San Francisco 49ers season. Retrieved May 20, 2016 from <https://en.wikipedia.org/wiki/2002_San_Francisco_49ers_season>

1. Historical game results (similar results for each team and years 1970-2015):

New Orleans Saints. Retrieved May 20, 2016 from <http://www.nfl.com/teams/neworleans%20saints/schedule?team=NO&season=2000&seasonType=REG>

1. Historical game information (similar results for each team and years 1970-2015):

New Orleans Saints @ Green Bay Packers. Retrieved May 20, 2016 from <http://www.nfl.com/gamecenter/2011090800/2011/REG1/saints@packers#menu=gameinfo&tab=recap>

1. Historical season information (similar results for seasons 1970-2015):

2000 NFL Season | The Football Database. Retrieved May 20, 2016 from <http://www.footballdb.com/seasons/nfl/2000>

1. Historical spread information:

Archived Closing NFL Odds, NFL Lines, NFL Point Spreads. Historical Pro Football: 2006-current. Retrieved May 20, 2016 from <http://www.footballlocks.com/archived_nfl_odds_lines_point_spreads.shtml>

Repole, Warren. Sunshine Forecast Downloadable Data Files. All Years from 1978 through 2013. Sunshine Forecast Enterprises. Retrieved may 20, 2016 from <http://www.repole.com/sun4cast/data.html#dataprior>

1. Larose, D. and Larose, C.D. 2014. *Discovering Knowledge in Data: An Introduction to Data Mining.* Wiley-Interscience.

**Completed Graduate Courses:**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Number*** | ***Course Name*** | ***Semester*** | ***Grade*** |
| COSC 578 | Database Management Systems I | Spring 2014 | A |
| COSC 650 | Computer Networks | Fall 2014 | A- |
| COSC 602 | Computer Vision and Image Processing | Fall 2014 | A |
| COSC 600 | Advanced Data Structure and Algorithms | Spring 2015 | B+ |
| COSC 609 | Software Project Management | Spring 2015 | A- |
| COSC 519 | Operating System Principles | Fall 2015 | A |
| COSC 647 | Application Software Security | Fall 2015 | A |
| COSC 757 | Data Mining | Spring 2016 | A |
| COSC 617 | Advanced Web Development | Spring 2016 | A |

**Projects Done in Other Courses:**

|  |  |
| --- | --- |
| **Course Number** | Project Description |
| COSC 757 | 1. *Violent Crime in Communities – analysis of dataset containing Federal Bureau of Investigation (FBI) Uniform Crime Report (UCR), Census, and LEMAS survey data relating to the community and law enforcement in order to understand how the indicators may predict and increased or decreased per capita crime levels.* |

**Project Requirements and Evaluation:**

1. Reading and presentation *AA%*

2. Implementation *BB%*

3. Examination *CC%*

4. Written report to the instructor *DD% (required regardless of % points)*

5. Public Presentation \* *EE% (required regardless of % points)*

\* The instructor/student will announce the presentation time and place to the Department faculty and students. Those interested in the topic may attend the presentation.

I, Mary J. Snyder propose to complete this project during the fall semester of 2016 and understand that this project and its derived materials (e.g., source code, written reports, presentation slides) are to reflect my own work, unless explicitly and appropriately referenced. Furthermore, I understand that plagiarism or other unattributed use of material not written by me is completely unacceptable, and **will be considered sufficient cause for a failing grade on the project.** For additional information on academic integrity policy at Towson University, I will visit [www.towson.edu/provost/resources/studentacademic.asp](http://www.towson.edu/provost/resources/studentacademic.asp) .

Student’s Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructor’s Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Graduate Program Director’s Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_