Course Syllabus

DATAMINING

Printed by: lisacabe

Program: Computer Science

1. Course number and name

CCPG1028 - DATAMINING

2. Credits and contact hours

3 credits and 3 contact hours

3. Instructor's course or coordinator's name

CARMEN KARINA VACA RUIZ

4. Text book, tittle, author, and year

• James, Gareth & James, Gareth & Witten, Daniela & Hastie, Trevor & Tibshirani, Robert. An introduction to statistical learningwith applications in R (2nd)

a.Other supplemental materials

• Hastie, Trevor & Tibshirani, Robert & Friedman, Jerome. The elements of statistical learning: data mining, inference and prediction (2nd)

• Marsland, S. Machine Learning: An Algorithmic Perspective (2nd)

5. Specific course information

a. Brief description of the content of the course (catalog description)

This course deals with the fundamental concepts of machine learning and predictive modeling. Supervised and unsupervised learning paradigms are presented including dimensionality reduction techniques. Evaluation metrics are presented and applied to real problems solved with the algorithms presented through the course. Recent applications of machine learning are discussed in areas such as finance, political science, human mobility among many others. Finally, readings and discussions about ethical dilemmas and trends in machine learning are conducted, considering the enormous adoption of these techniques in recent years.

b. This course is: Selected elective

6. Specific goals for the course

a. Specific outcomes of instruction

1.- Explain the differences between categories of machine learning paradigms by selecting the most appropriate one to solve a problem in industry or research.

2.- Implement supervised and unsupervised learning models for solving problems using data sets from different domains, interpreting the results obtained and communicating them to others.

3.- Evaluate the performance of an automatic learning model with a real data set

4.- Compare several supervised machine learning techniques using appropriate metrics

for the selection of the best performance algorithm.

5.- Explain the problem of overfitting and the available techniques for detection and management of such a problem.

6.- Distinguish ethical dilemmas that derive from the use of machine learning systems for decision-making in organizations.

b. Explicity indicate which of the student outcomes listed in Criterion 3 or any other outcomes are addressed by the course

7. Brief list of topics to be covered

1.- Introduction to machine learning

2.- Supervised learning: regression, classification, evaluation

3.- Bayesian classifiers

4.- Discriminative classifiers

5.- Model selection, feature selection, cross validation.

6.- Unsupervised learning

7.- Dimensionality reduction

8.- Trends in machine learning