**Bachelor of Science in Data Science and Business Administration**

**BSDSBA 2027**

**School Year 2023-2024**

**Course Outline**

|  |  |
| --- | --- |
| **School** | Aboitiz School of Information, Technology, and Entrepreneurship |
| **Course Code** | ACCS 000 |
| **Course Title** | Course Title |
| **Units** | 3.00 |
| **Term** | Choose Term |

|  |  |  |  |
| --- | --- | --- | --- |
| **Faculty Name** | Benjur Emmanuel Borja |  |  |
| **Email Address** | bborja@aim.edu |  |  |
| **Consultation Hours** | Tuesdays – 5:30 to 6:30 pm |  |  |

|  |  |
| --- | --- |
| **Program Staff** | Nicholas Paredes |
| **Email Address** | Nparedes@aim.edu |
| **Extension No.** |  |

**A. Course Description**

Machine Learning I is the first in the Machine Learning series. It covers topics on Introduction to Data Science and AI, KNN method, Linear Regressions and Regularization Methods, Support Vector Machine, Decision Trees, Random Forest, Gradient Boosting Methods.

At the end of the series, students will be able to:

1. Understand the modern age of big data and appreciate the new technologies, trends, and transformations brought about by big data generation
2. Learn and appreciate the processes involved in machine learning models from various datasets.
3. Understand various learning algorithms and how to apply them
4. Understand how to select appropriate machine learning models to various problems
5. Evaluate models generated from data
6. Implement the algorithms to a real problem and optimize the models learned
7. Develop the ability to evaluate the performance of machine learning algorithms and report on the expected accuracy in applying the models
8. Develop skills at presenting and communicating results obtained from machine learning models

Each course in the series contributes to the performance of these main objectives.

This course will introduce students to the world of data science and artificial intelligence, in particular, traditional machine learning models and predictive analytics. The course will start from conceptual understanding of models and algorithms and transition to hands-on applications of models to real-world big data in different fields. At the end of the course, students are expected to acquire the basic skills in implementing machine learning models and model evaluation.

*Note: Content of the course outline is subject to change.*

**B. Pre-requisites**

**C. Course Learning Outcomes  
At the end of the course, students will be able to:**

1. Describe the evolution of data science and artificial intelligence including the history and contemporary uses of machine learning;

2. Describe the challenge of generalization in machine learning models, especially the problems of overfitting and underfitting;

3. Conduct basic data cleaning and pre-processing of data in preparation for machine learning analysis;

4. Implement and evaluate k-Nearest Neighbors for classification and regression tasks;

5. Implement and evaluate linear models for classification and regression tasks, accordingly;

6. Implement and evaluate support vector machine, random forest, and gradient boosting models for classification and regression tasks;

**D. Course Contribution to Program Learning Goals and Objectives**

|  |  |
| --- | --- |
| **Course Learning Outcomes** | **Program Learning Goals/Objectives** |
| 1,2 | 1. BSDSBA students will be proficient in communicating business value.​ |
|  | 1.1 Visualize data with a focus on extracting actionable insights |
|  | 1.2 Convey compelling narratives and stories related to data and insights |
| 1 – 6 | 2. BSDSBA students will be adept in data science tools and platforms. |
|  | 2.1 Develop an in-depth understanding of data science and business analytics: data mining, machine learning, applied statistics, and predictive modeling. |
|  | 2.2 Apply the right data science principles in the analysis of business problems. |
|  | 2.3 Use cutting-edge tools and data mining software to analyze data and solve real-world problems. |
| 1 – 6 | 3. BSDSBA students will be critical and highly skilled in programming languages and big data analysis. |
|  | 3.1 Gain hands-on experience with programming languages and big data tools through coursework, research, and engagements with actual field practitioners. |
|  | 3.2 Recognize and analyze ethical issues in business situations and make well-reasoned ethical management decisions |
| 1 – 3 | 4. BSDSBA students will be informed decision-makers and implementors who can work with teams. |
|  | 4.1 Practice real-world problem analysis to guide pitching and decision-making. |
|  | 4.2 Demonstrate knowledge of algorithms to build machine intelligence utilized in business decision-making. ​ |
|  | 4.3 Demonstrate use of teamwork, leadership skills, and decision-making. |

**E. Learning Methodology**

The lecture and hands-on activities will be supplemented with:

* PPT slides
* Jupyter Notebooks
* Exam
* Real world data sets
* *Recitation class will be a venue for in-depth quizzes, discussions, and individual and/or learning team coaching.*

**F. Grading Criteria**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | |  | **Weight** |
| Recit Classes, Homeworks, and Quizzes |  | | (Group) | **20%** |
| Class Participation |  | | (Individual) | **10%** |
| 2 Long Exams (Midterm and Finals) |  | | (Individual) | **50%** |
| Final Project |  | | (Group) | **20%** |
|  |  | |  |  |
| **Total** |  |  | | **100%** |

*\*Final Grades automatically calculated in the ALICE Grade Center are not conclusive and are subject to Program Deliberations of the Faculty at the end of the Program Term*

**G. Student Responsibilities and Conduct**Students are expected to conduct themselves with the utmost professionalism in all classes. Information and policies on student responsibilities and conduct, including dysfunctional behavior (such as attendance, plagiarism, cheating, etc.) and grievance procedures are in the Student Handbook.

**H. Course Schedule**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Session Number | Session Topic (in-session) | Pre-session Activity | *Faculty/Resource Speaker* | Learning or Case Materials | Supplementary Readings | Assessment activities | Requirements or Submissions | Post-session Activity |
| 1)  January 13, 2025 11:00 AM -12:30 PM | Class management and course introduction  Introduction to data science | Discuss the course outline and how the class will be conducted using a combination of lectures, supporting laboratory exercises, cases and Python notebooks.  Introduce the three ML courses and the expectations in each of the courses.  Discussion from data to insights, and on rise of big data and digital transformation. |  | Students must be comfortable with markdowns and editing Jupyter notebooks/Google Colabs. |  |  |  |  |
| 2)  January 16, 2025 11:00 AM -12:30 PM | Mathematics foundations for data science | Review of matrix algebra and calculus (derivatives). |  |  |  |  |  |  |
| 3)  January 20, 2025 11:00 AM -12:30 PM | History of learning systems  General usage and applications of machine learning and its limitations | Definition of learning systems. Goals and applications of machine learning.  Designing a learning system including training data, function approximation, and approximation. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 4)  January 23, 2025 11:00 AM -12:30 PM | Preliminary examples: Discuss simple use cases and solutions | Understanding data (statistics)  Measuring Success: Training and Test Data, Metrics of predictive accuracy.  Formulating problems, designing the analytics, designing the implementation |  |  |  |  |  |  |
| 5)  January 27, 2025 11:00 AM -12:30 PM | Overview of k-Nearest Neighbors (kNN) classifier and Regressor | Introduction to the concept of KNN including assumptions and formulation. Intuition building exercise. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  | Quiz |  |  |
| 6)  January 30, 2025 11:00 AM -12:30 PM | KNN: Classifier | Hands-on exercises of KNN using simple data to real-world data.  Discussion and interpretation of results.  Illustrations/Cases:  1. Flower species classification;  2. Prediction on whether a tumor is malignant or not based on tissue measurements using Boston cancer data; |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 1: Jupyter Notebook, due 1 week after |  |
| 7)  February 3, 2025 11:00 AM -12:30 PM | KNN: Regressor | Hands-on exercises of KNN using simple data to real-world data.  Discussion and interpretation of results. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 2: Jupyter Notebook, due 1 week after |  |
| 8)  February 6, 2025 11:00 AM -12:30 PM | Recitation Class 1 on KNN Classifier and Regressor | Hands-on modeling of the topics  Discussion of answers |  |  |  | Quiz |  |  |
| 9)  February 10, 2025 11:00 AM -12:30 PM | Linear Regression | Introduction to the concept of linear regression – OLS and regularization concepts and conduct of simple hands-on exercises in Excel and in Jupyter notebook.  Discussion of fitting parameters and cost-function. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 10)  February 13, 2025 11:00 AM -12:30 PM | Linear Regression | Hands-on exercises of linear regression problems with real-world data.  Illustration/Cases:  1. Boston House Pricing Data  Interpretation of the results of linear regression. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 11)  February 17, 2025 11:00 AM -12:30 PM | Regularization: L1 and L2 types | Demonstration on how generalization of prediction can be improved by including regularization in the error formulation. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 3: Jupyter Notebook, due 1 week after |  |
| 12)  February 20, 2025 11:00 AM -12:30 PM | Recitation Class 2 on Linear Regression and regularization | Hands-on modeling  Discussion of answers |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  | Quiz |  |  |
| 13)  February 24, 2025 11:00 AM -12:30 PM | **Midterm Exam** |  |  |  |  |  |  |  |
| 14)  February 27, 2025 11:00 AM -12:30 PM | Linear Classifiers: Logistic Regression | Introduction to the concept of logistic regression as classifier and conduct of simple hands-on exercises in Jupyter notebook.  Cases/Examples:   1. Forge data set;   Revisiting the Boston cancer data set.  Discussion of results and its interpretation/s. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 15)  March 3, 2025 11:00 AM -12:30 PM | Support Vector Machine | Cases/Examples:   1. Forge data set;   Revisiting the Boston cancer data set. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 4: Jupyter Notebook, due 1 week after |  |
| 16)  March 6, 2025 11:00 AM -12:30 PM | Insights on Linear Models for multiclass classification | Cases/Examples:   1. Synthetic Blobs   We use a two-dimensional dataset, where each class is given by data sampled from a Gaussian distribution. |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 17)  March 10, 2025 11:00 AM -12:30 PM | Recitation Class 3 on Linear Classifiers | Hands-on modeling  Discussion of answers |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  | Quiz |  |  |
| 18)  March 13, 2025 11:00 AM -12:30 PM | **Basic approaches in data cleaning and preprocessing** | Case illustration: Bank marketing data |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 19)  March 17, 2025 11:00 AM -12:30 PM | Decision Trees: Classifier | Introduction to the concept of decision trees as classifier  Cases/Example:   1. Boston Cancer data   Kaggle Titanic dataset |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 20)  March 20, 2025 11:00 AM -12:30 PM | Decision Trees: Regressor | Introduction to the concept of decision trees as regressor  Cases/Example:  House price dataset |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 21)  April 7, 2025 9:00AM – 10:30 AM | Random Forest: Classifier | Introduction to the concept of decision trees as classifier  Cases/Example:   1. Boston Cancer data   Kaggle Titanic dataset |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 5: Jupyter Notebook, due 1 week after |  |
| 22)  April 7, 2025 11:00 AM -12:30 PM | Random Forest: Regressor | Introduction to the concept of random forest as regressor  Cases/Example:  House price dataset |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 23)  April 10, 2025 9:00 AM – 10:30 AM | Recitation Class 4 on Decision Trees and Random Forest | Hands-on modeling  Discussion of answers |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  | Quiz |  |  |
| 24)  April 10, 2025 11:00 AM -12:30 PM | Gradient-Boosting Machine: Classifier | Introduction to the concept of GBM as classifier  Cases/Example:   1. Boston Cancer data   Kaggle Titanic dataset |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 25)  April 14, 2025 9:00 AM – 10:30 AM | Gradient-Boosting Machine: Regressor | Introduction to the concept of GBM as regressor  Cases/Example:  House price data |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  | Homework 6: Jupyter Notebook, due 1 week after |  |
| 26)  April 14, 2025 11:00 AM -12:30 PM | GBM: Other modifications | Introduction to XGBoost and Adaboost  Hands-on implementation on simple data |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  |  |  |  |
| 27)  April 21, 2025 11:00 AM -12:30 PM | Recitation Class 5 on Gradient Boosting Machine | Hands-on modeling  Discussion of answers |  | Class lessons are guided by Jupyter notebooks prepared and the main reference. |  | Quiz |  |  |
| 28)  April 24, 2025 11:00 AM -12:30 PM | **Final Exam** |  |  |  |  |  |  |  |
| 29)  April 28, 2025 11:00 AM -12:30 PM | Final Project Presentation |  |  |  |  |  |  |  |
| 30)  May 5, 2025 11:00 AM -12:30 PM | Final Project Presentation |  |  |  |  |  |  |  |

**Required References**

Most of the materials in this course are written, conceptualized and implemented by Prof. Benjur Emmanuel L. Borja. None of the contents can be reproduced, distributed or patented without written permission.

**Additional References**

Mitchell T, Machine Learning, McGraw-Hill, 1997

S. Rogers and M. Girolami, A first course in Machine Learning, CRC Press, 2011

C. Bishop, Pattern Recognition and Machine Learning, 2007

D. Barber, Bayesian Reasoning and Machine Learning, 2012

Duda, Hart and Stork, Pattern Classification, Wiley-Interscience.