Course guide
230337 - AIPRAC - Problem Solving with Artificial Intelligence: a Practical Approach

Last modified: 17/06/2024

Unit in charge: Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications.

Degree: BACHELOR'S DEGREE IN TELECOMMUNICATIONS TECHNOLOGIES AND SERVICES ENGINEERING (Syllabus 2015). (Optional subject).
BACHELOR'S DEGREE IN ELECTRONIC ENGINEERING AND TELECOMMUNICATION (Syllabus 2018). (Optional subject).

Academic year: 2024 **ECTS Credits:** 2.0 **Languages:** Catalan

LECTURER

Coordinating lecturer: Enric Monte Moreno

Others:

PRIOR SKILLS

Python programming and good knowledge of the subjects of "Vector Calculus", "Signals and Systems" and "Probability and Statistics"

REQUIREMENTS

"Vector Calculus" and "Probability and Statistics"

TEACHING METHODOLOGY

One hour of each class session will be dedicated to explaining the database that will be used, and its statistical and geometric properties. Then, considering the properties of the database, the choice of an appropriate artificial intelligence method for it will be justified. The next two hours will be dedicated to working with the database, using a Jupyter 'notebook' in Python language.

LEARNING OBJECTIVES OF THE SUBJECT

The objective of the course is to provide students with tools to analyze the properties of a database related to a real application, and to decide which artificial intelligence technique is most appropriate to solve the problem.

STUDY LOAD

Type	Hours	Percentage
Hours small group	14,0	28.00
Hours large group	6,0	12.00
Self study	30,0	60.00

Total learning time: 50 h

CONTENTS

Exploratory Data Analysis (EDA) and Logistic Regression

Description:
• Database: South African Heart Disease
• Technique: EDA and Logistic Regression

Specific objectives:
Analyze the South African Heart Disease dataset, which contains information on male patients, including characteristics like age, blood pressure, cholesterol levels, smoking history, and presence of heart disease. Utilize EDA techniques (histograms, scatterplots, boxplots) to understand the distribution of these features and explore potential relationships between them and heart disease. Apply logistic regression to predict the likelihood of a patient having heart disease based on these features. Evaluate the model performance using metrics like accuracy, precision, and recall. Analyze the coefficients of the model to understand which factors have the most significant influence on heart disease risk.

Related activities:
Jypiter Notebook in python

Full-or-part-time: 3h
Practical classes: 3h

Classification with K-Nearest Neighbors (KNN) and Decision Trees

Description:
• Database: Customer churn prediction
• Technique: K-Nearest Neighbors & Decision Trees

Specific objectives:
Analyze a database containing customer information. Use this data to predict which customers are likely to churn (stop using the service). Apply KNN and decision trees to classify customers as likely churners or not. Compare the performance of both models and analyze the factors influencing churn.

Related activities:
Jypiter Notebook in python

Full-or-part-time: 3h
Practical classes: 3h

Dimensionality Reduction with Principal Component Analysis (PCA)

Description:
• Database: Stock market data
• Technique: Principal Component Analysis (PCA)

Specific objectives:
Analyze a database of historical stock prices for several companies. Use PCA to reduce the dimensionality of the data while preserving most of the information. Visualize the reduced data and interpret the principal components. Analyze how these components relate to stock price movements.

Related activities:
Jypiter Notebook in python

Full-or-part-time: 3h
Practical classes: 3h

Clustering with K-Means

Description:
• Database: Customer segmentation
• Technique: K-Means Clustering

Specific objectives:
Analyze a database of historical stock prices for several companies. Use PCA to reduce the dimensionality of the data while preserving most of the information. Visualize the reduced data and interpret the principal components. Analyze how these components relate to stock price movements

Related activities:
Jupyter notebook in python.

Full-or-part-time: 3h
Practical classes: 3h

Time Series Forecasting with ARIMA

Description:
• Database: Daily temperature data
• Technique: Autoregressive Integrated Moving Average (ARIMA)

Specific objectives:
Analyze a database of daily temperature data. Use ARIMA models to forecast future temperatures. Evaluate the forecasting accuracy using metrics like mean squared error. Analyze seasonal trends and patterns in the data and incorporate them into the model if necessary.

Related activities:
Jupyter notebook in python.

Full-or-part-time: 3h
Practical classes: 3h

Text Classification with Support Vector Machines (SVM)

Description:
Database: Movie reviews
Technique: Support Vector Machines (SVM)

Specific objectives:
Analyze a database of movie reviews. Preprocess the text data (e.g., remove stop words, perform stemming/lemmatization). Use SVMs to classify movie reviews as positive, negative, or neutral. Explore different text processing techniques and evaluate their impact on model performance.

Related activities:
Jupyter notebook in python

Full-or-part-time: 3h
Practical classes: 3h

Final Session

Description:
Recap of what has been learned

Full-or-part-time: 2h
Practical classes: 2h

GRADING SYSTEM

The grading will be based on the 'notebook' submitted at the end of each work session.

BIBLIOGRAPHY

Basic:
- James, G., Witten, D., Hastie, T., & Tibshirani, R.. An introduction to statistical learning [on line]. New York: springer, 2013 [Consultation: 12/06/2024]. Available on: https://www.statlearning.com/.
- Richert, Willi.. Building Machine Learning Systems with Python [on line]. 2013. New York: Packt Publishing, 2013 [Consultation: 17/06/2024]. Available on: https://www.packtpub.com/product/building-machine-learning-systems-with-python/9781782161400.