WATCH

Watch the project overview video and then review the notes below to get started.

**Project 5 - Classification Overview Video**

## EXTRA TIPS

For this project you will be using the pd.get\_dummies function, this function can be a little hard to understand, watch this video to get a better grasp of it and why it is useful.

**Project 5 - Classification Pandas get\_dummies Overview Video**

REVIEW NOTES

This project will give you practice running three different classification algorithms (k-Nearest Neighbors, Support Vector Machines, and Logistic Regression) and comparing the results.

Use the [**project 5 notebook**](https://elearning.mines.edu/courses/52392/files/5459027/download?wrap=1)[Download project 5 notebook](https://elearning.mines.edu/courses/52392/files/5459027/download?download_frd=1) and the [**titanic-data.csv**](https://elearning.mines.edu/courses/52392/files/5458983/download?wrap=1)[Download titanic-data.csv](https://elearning.mines.edu/courses/52392/files/5458983/download?download_frd=1)file you will be instructed to import the data file within your jupyter notebook, and answer some questions about the best performing model from your experimentation.

Remember, there are great resources in addition to our discussions, including this [Python reference tutorialLinks to an external site.](https://docs.python.org/3/tutorial/index.html).

SUBMITTING YOUR SOLUTION

To submit your notebook, first clear all the cells (this won't matter too much this time, but for larger data sets in the future, it will make the file smaller). Then use the File->Download As->Notebook to obtain the notebook file. Finally, submit the notebook file with your solutions to Canvas.

View Rubric

| Project 5 | | |
| --- | --- | --- |
|  |  |  |
| **Project 5** | | |
| **Criteria** | **Ratings** | **Pts** |
| **Problem 1**  **view longer description** | 35 pts  Full Marks  Questions 1-7 answered, (ie. file is loaded, data exploration is performed, questions 3-7 have a graph or written answer backed by data)  0 pts  No Marks | / 35 pts |
| **Problem 2**  **view longer description** | 10 pts  Full Marks  Data transformation is correctly performed  0 pts  No Marks | / 10 pts |
| **Problem 3**  **view longer description** | 30 pts  Full Marks  Inputs/targets correctly defined, data is split, classifier with k=5 is created, model use of fit(), predict(), and score  0 pts  No Marks | / 30 pts |
| **Problem 4**  **view longer description** | 15 pts  Full Marks  Scorer variable is created and used, cross validation score is calculated, confusion matrix and classification report are created.  0 pts  No Marks | / 15 pts |
| **Problem 5**  **view longer description** | 15 pts  Full Marks  SVM model is created and used, cross validation score is calculated, confusion matrix and classification report are created.  0 pts  No Marks | / 15 pts |
| **Problem 6**  **view longer description** | 15 pts  Full Marks  Logistic Regression model created, then scored using f-score and confusion matrix  0 pts  No Marks | / 15 pts |
| **Problem 7**  **view longer description** | 5 pts  Full Marks  Thoughtful and sensible answers to discussion questions.  0 pts  No Marks | / 5 pts |
| Total Points: 0 | | |