

Exam



HONOR PLEDGE

As a student at CSU, Chico, I hold myself to a high standard of integrity, and by signing/accepting the statement below I reaffirm my pledge to act ethically by honoring the CSU, Chico Academic Integrity Policy. I will also encourage other students to avoid academic misconduct.

I acknowledge that the work I submit is my individual effort. I did not consult with or receive any help from any person or other source. I also did not provide help to others. I may work with others only if the instructor gave specific instructions, and only to the extent allowed by the instructor.

I understand that suspected misconduct on this assignment/exam will be reported to the Office of Student Conduct, Rights, and Responsibilities and, if established, will result in disciplinary sanctions up through Dismissal from the University and a grade penalty up to a grade of “F” for the course.

I understand that by choosing to access this exam I am acknowledging this full statement and holding it to be true. If I fail to comply, the instructor will not grade this work and may assign a grade of “0” or “F” for the exam.



EXAM Q&A

Selecting this link will redirect you to the Discussion Board - Exam forum



EXAM Preparation

Attached Files: [titanic_train.csv](#) (58.889 KB)

*** THIS IS AN EXAM and ALL WORK SHOULD BE DONE INDEPENDENTLY ***

You are expected to create your own notebook, performing your own research and investigations to support your efforts along the way. You may ask questions for clarification of the notebook on the Discussion Board, but you may not share solutions or work together.

You will have 1 week to prepare your notebook. On the day of the exam you will be given a worksheet that will guide you through exploration of the data and predictions. You are to rely on your notebook for answers. You will complete the worksheet and upload it along with your **Jupyter Notebook (.ipynb file)** to the Exam submission link.

We will be using the Kaggle **titanic** dataset for our Exam.

PREPARING YOUR NOTEBOOK:

You should prepare your personal, unique, custom **Titanic notebook** in advance, and complete it to the best of your ability.

- You may perform independent research and incorporate discoveries into your notebook solution, but only those you understand, can explain, and can modify/manipulate. In other words, your notebook should be your own.
- You will be required to load the data. I suggest storing it in a pandas Dataframe.
- You will want to apply methods to visualize, interpret, and explore the data.
- You will want to apply methods for pre-processing/engineering the data. This may involve ignoring, removing, and/or replacing values, columns, or rows. I suggest you review data transformations such as scaling and encoding, research and explore *mean imputation for missing data*.
- You will be required to create multiple **SKLearn models** for binary classification of unseen data. I suggest you revisit *logistic regression, SVC, naive bayes, decision trees, random forests, and mlps*.
- You will be required to evaluate the performance of your models. I suggest you revisit appropriate metrics and explore options for visualations (e.g., matplotlib and seaborn graphs and plots).
- **KEEP A BACKUP COPY OF YOUR NOTEBOOK ALONG THE WAY** - a 'lost' notebook will not be grounds for an exemption or extension.

YOUR GOAL:

Prepare a notebook that is thorough and well-documented and aims at providing a comparison of models, each performing as best as possible on data that you have prepared and pre-processed to the best of your ability. **Your ML models will be limited to those supported by SciKit-Learn.**

FINAL EXAM DAY:

- You will be asked to rely on **your notebook** for guided exploration, to generate responses to questions I will pose in a worksheet. You may need to make modifications to your methods or models to properly address my questions. Therefore, it is imperative that you understand and document your code well, as you create your notebook during the preparatory week.
- **Your original notebook must be uploaded the evening before the exam.** (Attempt 1)
- **You *must* be and remain present for the duration of the Exam Period (50 min),** which is the only time in which you may ask me questions about the exam questionnaire.
- **You will have until the end of the exam evening (PT) to complete the worksheet.**
- **Your updated notebook and completed worksheet must be uploaded the evening of the exam.** (Attempt 2)



TITANIC Notebook - Tips#1

Attached Files:  EXAM_TITANIC_GettingStarted.ipynb (1.912 KB)

DATAFRAMES --- Get Better Acquainted with available methods & attributes of this Pandas data structure. Some of particular interest/use may include:

- [Tutorial](#)
- `apply`
- `count`
- `describe`
- `drop`
- `dropna`
- `head`
- `info`
- `isnull`
- `max`
- `mean`
- `mode`
- `unique`

VISUALIZATIONS

[SEABORN](#)

- [tutorials](#)
- [Heatmap](#)
- [Countplot](#)
- [Displot](#)

[MATPLOTLIB](#)

- [tutorials](#)
- `figure`
- `title`
- *etc.*



TITANIC Notebook - Tips#2

Attached Files:  EXAM_TITANIC_NextSteps.ipynb (2.398 KB)

MODELING (classifiers)

- [Logistic Regression](#)
- [Linear Support Vector Classifiers](#)
- [Gaussian Naive Bayes](#)
- [Decision Trees](#)
- [Random Forests](#)
- [Multi-Layer Perceptrons](#)

EVALUATION & SELECTION

Common HYPERPARAMETERS, ATTRIBUTES, and METHODS

- `shuffle data`
- `random_state`
- `max_iter`
- *C or alpha*
- *L1 or L2 penalty*
- `solver`
- `fit`
- `predict`
- `score`
- TREES
 - `n_estimators`
 - `max_depth`; `get_depth`
 - `min_samples_leaf`; `get_n_leaves`
 - `min_samples_split`
 - `max_features`
 - `max_leaf_nodes`
- MLPs
 - `hidden_layer_sizes`
 - `activation`
 - `solver`
 - `alpha`
 - `batch_size`
 - `learning_rate`
 - `max_iter`
 - `shuffle`
 - `random_state`
 - `early_stopping`
 - `momentum`
 - `coefs_`
 - `intercepts_`
 - `n_iter`
 - `n_layers`
 - `n_outputs`

HELPFUL TECHNIQUES

- Transformers
 - `scalers`
 - `encoders`
- Cross Validation
- Grid Search