EE660 Project Title

Author(s), email contacts

Date

# Project Homepage (mandatory)

**You must report the homepage of your project Git repository. You can include details of code structure or miscellaneous information in a ReadMe file in the repository.**

# Abstract

A brief, informative description of your project. Include the problem, approach, and key results.

# Problem Statement and Goals

A brief description of the problem that you are trying to solve, and/or the goals you want to achieve. Explain why it’s important or interesting, and why it’s not trivial.

Normally difficulty can come from any of the following sources:

1. A physical model that’s inherently complicated and hard to abstract.
2. High dimensionality of feature space.
3. Sparsity.
4. Nonlinear behaviors.
5. Limited number of training samples.
6. Significant amounts of preprocessing required.

# Literature Review (if applicable)

Briefly describe existing approaches to your problem. The literature review doesn’t need to be exhaustive, but it should cover well known publications so you are aware of their approach, tools and results.

# Prior and Related Work (Mandatory)

If this project is an extension of some work you previously did or are currently doing outside of EE 660, briefly summarize this other work, and clearly distinguish it from your EE 660 project work.

If you have no prior or related work, state so (*e.g.*, this can be done as a single heading: “Prior and Related Work - None”).

# Project Formulation and Setup

In here, please describe your model and your algorithm in detail (with formulas and flowcharts). No need to include details in training or testing process in this section, but do include details of your model, such as the knowns and unknowns, parameters to tune, etc..

If you are implementing an existing approach, explain why it fits best in your case, also explain in detail what you did to adapt the method to your case, and explain the assumptions or “tricks” you used.

If you are developing an original idea, please clearly state so, explain your algorithm in detail, and clearly point out your point of innovation and any advantages of your approach.

# Methodology

Describe the entire procedure you followed to get your model, train it, and evaluate it. This will include (to be stated in the sequence used): the hypothesis set (or you may even have multiple hypothesis sets at different levels), at what point the dataset was brought into the picture, when its various subsets were used (validation or test sets, etc.), training procedure, model selection (if any), validation and test procedures.

You might consider using a flow chart to illustrate your framework. Details in each step could be relatively brief, as they can be covered in the next section.

# Implementation

Report your implementation details and results here. You can follow the steps described in Section 6 “Methodology”.

Typically the following steps are covered:

## Feature Space

Describe the dataset you used, explain the meaning and data type (integer, real, string or binary, categorical, etc.) of each feature.

## Pre-processing of Training Data

Main dataset consists of different data sets, which are: Training Set, Testing Set, and Additional Set. Training set includes

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Description | Usable |  |
| ID | The id of the record | Yes |  |
| Address | The Full Address of Trap | No |
| Date | Date the test is performed | Yes |  |
| Speicies | The species of mosquitos | No |  |
| Trap ID | The id of the trap | Yes |  |
| NumMosquitos | The number of the mosquitos | Yes |  |
| Latitude Longitutde | Geography data | Yes |  |
| Block Street Address address accuracy | Social location description | No |  |
| WNVpresented | Whether the virus presents | yes |  |
| Weather data and spray location | Additional data | no |  |

While in Test set, for consideration from the Website, the mosquito number is not available, and thus became a secondary prediction target.

Additional Data Set is for enhancing accuracy, including spray data and weather data.

### Feature Extraction

Locations features in this dataset are extremely duplicated. We only use Latitude, Longitude, block number and Trap Number for location features. Replace the mosquito species with number from 1 to 8. Change Date to Uniform Time Value

|  |  |
| --- | --- |
| Type | Feature Value |
| CULEX PIPIENS/RESTUANS | 1 |
| CULEX RESTUANS | 2 |
| CULEX PIPIENS | 3 |
| CULEX SALINARIUS | 4 |
| CULEX TERRITANS | 5 |
| CULEX TARSALIS | 6 |
| UNSPECIFIED CULEX | 7 |
| CULEX ERRATICUS | 8 |

### 

## Training Process

Describe how you train your model, the classifiers or regression processes you use, and the parameters you chose.

If a parameter is chosen by heuristics, state so. If a parameter is chosen by some model selection or validation process, you can cover it in the next subsection.

Analyze the complexity of your hypothesis set. Give the number of samples you have and the dimension of the pre-processed feature space. Explain what you did to avoid overfitting and underfitting.

## Testing, Validation and Model Selection

You must run enough tests to show that your method works. Perform cross validation if needed.

If you have multiple potential models in the beginning, explain how you perform model selection in detail.

Report testing results here. In the case of classification, you can choose two salient features and plot your decision boundaries w.r.t. them. In the case of regression, you can plot the resulting regression function in 3D (as a function of two salient features) or in a few 2D plots (each as a function of one salient feature).

If you have a large dataset, it might be reasonable to use validation process to tune your model. If you did that, state so.

# Final Results

Describe and document your final results. Different from the results in Section 7.4., the results here should be the final performance of your system(s) and an estimate of its out of sample performance. Figures, plots, and/or tables can be useful.

**If you are working on an online competition, report the performance of your best submission and compare it to others on the leader board.** If you want to compare your results with other work, do so here or in the Interpretation section below.

# Interpretation

Why do you think the results came out the way they did? What has been learned from them? Anything particularly noteworthy or unexpected?

# Summary and conclusions

Briefly summarize key findings, and optionally state what would be interesting or useful to do next.

# Reference

Never forget to cite your references.