Master’s Project Outlines

**Timeline**

2/14 – Try out different algorithms and start evaluations of algorithms.

2/29 – Need to have all implementations done and evaluated.

3/21 – First draft of the write up done and reviewed by committees.

4/4 – Final draft done and submitted.

5/27 - Final deadline for File Thesis with Graduate Studies.

**Evaluation**

Manual labeling

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 57** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.04925 | 0.00067 | 0.00001 | 0.00001 | 0.01128 | 0.00015 |
| EST Score Filter | 0.29542 | 0.00174 | 0.00006 | 0.00002 | 0.06766 | 0.00048 |
| NBC | 0.04218 | 0.00076 | 0.00178 | 0.00028 | 0.01103 | 0.00021 |
| Modified NBC | 0.04446 | 0.00085 | 0.00030 | 0.00006 | 0.01041 | 0.00018 |
| Decision Tree | 0.00873 | 0.00075 | 0.32312 | 0.23560 | 0.25116 | 0.18156 |
| Random Forests | 0.00940 | 0.00078 | 0.00123 | 0.00007 | 0.00310 | 0.00008 |
| SVM | 0.06633 | 0.01269 | 0.06153 | 0.08859 | 0.06265 | 0.06568 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 60** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.25081 | 0.00279 | 0.00000 | 0.00000 | 0.00785 | 0.00009 |
| EST Score Filter | 0.17012 | 0.00204 | 0.00050 | 0.00002 | 0.00580 | 0.00005 |
| NBC | 0.16115 | 0.02116 | 0.00747 | 0.00908 | 0.01228 | 0.00813 |
| Modified NBC | 0.17066 | 0.00206 | 0.00033 | 0.00005 | 0.00566 | 0.00006 |
| Decision Tree | 0.14406 | 0.00708 | 0.42270 | 0.14028 | 0.41400 | 0.13590 |
| Random Forests | 0.15113 | 0.00171 | 0.00064 | 0.00004 | 0.00535 | 0.00006 |
| SVM | 1.00000 | 0.00000 | 0.00000 | 0.00000 | 0.03128 | 0.00022 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 61** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.29954 | 0.01224 | 0.00000 | 0.00000 | 0.13223 | 0.00526 |
| EST Score Filter | 0.92182 | 0.00679 | 0.00243 | 0.00143 | 0.40834 | 0.00658 |
| NBC | 0.14974 | 0.00909 | 0.13115 | 0.01862 | 0.13940 | 0.00987 |
| Modified NBC | 0.28355 | 0.01183 | 0.00235 | 0.00103 | 0.12650 | 0.00556 |
| Decision Tree | 0.09597 | 0.01542 | 0.65953 | 0.09608 | 0.41059 | 0.05007 |
| Random Forests | 0.14561 | 0.00873 | 0.06876 | 0.00664 | 0.10270 | 0.00506 |
| SVM | 0.37152 | 0.00910 | 0.03955 | 0.02912 | 0.18625 | 0.01692 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 62** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.39254 | 0.00604 | 0.00000 | 0.00000 | 0.22867 | 0.00414 |
| EST Score Filter | 0.98687 | 0.00188 | 0.00060 | 0.00057 | 0.57516 | 0.00626 |
| NBC | 0.04305 | 0.00519 | 0.08417 | 0.01214 | 0.06018 | 0.00562 |
| Modified NBC | 0.07441 | 0.00611 | 0.00344 | 0.00525 | 0.04475 | 0.00424 |
| Decision Tree | 0.02704 | 0.00941 | 0.59759 | 0.18252 | 0.26545 | 0.07728 |
| Random Forests | 0.04079 | 0.00500 | 0.03350 | 0.00269 | 0.03776 | 0.00298 |
| SVM | 0.17537 | 0.07043 | 0.05139 | 0.02368 | 0.12347 | 0.03699 |

Likelihood > 0.7

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 57** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.45021 | 0.00220 | 0.00000 | 0.00000 | 0.17820 | 0.00091 |
| EST Score Filter | 0.59276 | 0.00124 | 0.00019 | 0.00003 | 0.23475 | 0.00071 |
| NBC | 0.43427 | 0.00219 | 0.00333 | 0.00023 | 0.17390 | 0.00089 |
| Modified NBC | 0.44659 | 0.00217 | 0.00091 | 0.00008 | 0.17732 | 0.00091 |
| Decision Tree | 0.03101 | 0.00067 | 0.45433 | 0.27024 | 0.28681 | 0.16340 |
| Random Forests | 0.01982 | 0.00066 | 0.02081 | 0.00031 | 0.02042 | 0.00036 |
| SVM | 0.35620 | 0.10955 | 0.05207 | 0.06065 | 0.17245 | 0.01601 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 60** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.85769 | 0.00108 | 0.00006 | 0.00001 | 0.14098 | 0.00012 |
| EST Score Filter | 0.84252 | 0.00098 | 0.00068 | 0.00003 | 0.13900 | 0.00013 |
| NBC | 0.03302 | 0.00234 | 0.03235 | 0.00944 | 0.03246 | 0.00791 |
| Modified NBC | 0.08985 | 0.00356 | 0.00343 | 0.00041 | 0.01763 | 0.00067 |
| Decision Tree | 0.00540 | 0.00064 | 0.27986 | 0.07012 | 0.23476 | 0.05855 |
| Random Forests | 0.00653 | 0.00028 | 0.00083 | 0.00004 | 0.00177 | 0.00006 |
| SVM | 0.18109 | 0.00146 | 0.00000 | 0.00000 | 0.02976 | 0.00029 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 61** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.35446 | 0.01603 | 0.00042 | 0.00042 | 0.17001 | 0.00904 |
| EST Score Filter | 0.92793 | 0.00633 | 0.00268 | 0.00153 | 0.44579 | 0.00945 |
| NBC | 0.20108 | 0.01541 | 0.18337 | 0.01662 | 0.19180 | 0.00939 |
| Modified NBC | 0.34285 | 0.01726 | 0.00344 | 0.00249 | 0.16603 | 0.00907 |
| Decision Tree | 0.09007 | 0.01366 | 0.44294 | 0.08030 | 0.27398 | 0.04140 |
| Random Forests | 0.15348 | 0.01141 | 0.10311 | 0.00928 | 0.12728 | 0.00870 |
| SVM | 0.37109 | 0.01147 | 0.03150 | 0.01137 | 0.19407 | 0.00850 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Deployment 62** | FP Rate | | FN Rate | | Overall Error Rate | |
| Method | Mean | SD | Mean | SD | Mean | SD |
| Bandwidth Filter | 0.54189 | 0.00513 | 0.00561 | 0.00331 | 0.41839 | 0.00553 |
| EST Score Filter | 0.99046 | 0.00169 | 0.00248 | 0.00172 | 0.76292 | 0.00782 |
| NBC | 0.07243 | 0.00778 | 0.19488 | 0.01750 | 0.10065 | 0.00847 |
| Modified NBC | 0.13276 | 0.00789 | 0.00426 | 0.00337 | 0.10320 | 0.00688 |
| Decision Tree | 0.03000 | 0.00483 | 0.55081 | 0.10170 | 0.15010 | 0.02336 |
| Random Forests | 0.04694 | 0.00587 | 0.13700 | 0.01033 | 0.06771 | 0.00478 |
| SVM | 0.11067 | 0.01521 | 0.10279 | 0.04786 | 0.10891 | 0.00531 |

**The format of the paper:**

Inntruction/ background – talking about the qraat project and the problem descriptions.

Data: describing how the data was collected and labeled.

Method: describing the implementation of the 3 classifiers and its evulations.

Conclusion: Which method is better and why we choose to use it.

**Tables:**

* est\_bearing – likelihood of the expected bearing.
* est\_class – manually labeling of the data with differnet threshold on different parameters.
* est\_class2 – labeling of the data with likelihood > 0.7 as pulse.
* classifier\_performance – the count of TP, TN, FP, FN… etc for manual labeling.
* Classifier\_performance2 – the count of TP, TN, FP, FN… etc for likelihood labeling.
* estscore2 – the scoring of the est using the data driven est z value scoring method.
* est\_mean\_and\_var – the mean and variance of each data for each site and deployment for NBC.
* est\_mean\_and\_var2 – the mean and variance of each data for each site and deployment for NBC for second set of data.
* probability\_of\_discrete\_data – the information needed for the mixed probability distrubtoin in the modified BC.
* probability\_of\_discrete\_data2 – the information needed for the mixed probability distrubtoin in the modified BC for second set of data.
* Decision\_tree – the trees for each deployment and site combinations.
* Random\_forests – the forests for each deployment and site combinations.

**Database Schema:**

CREATE TABLE IF NOT EXISTS qraat.`classifier\_performance` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`validation` int(10) unsigned NOT NULL,

`TP` int(10) unsigned NOT NULL,

`TN` int(10) unsigned NOT NULL,

`FP` int(10) unsigned NOT NULL,

`FN` int(10) unsigned NOT NULL,

`total\_records` int(10) unsigned NOT NULL,

`classifier\_type` text NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`probability\_of\_discrete\_data` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`data\_type` text NOT NULL,

`data\_value` double unsigned NOT NULL,

`probability` double unsigned NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`est\_mean\_and\_var` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`probability` double unsigned NOT NULL,

`band3\_mean` double unsigned NOT NULL,

`band3\_var` double unsigned NOT NULL,

`band10\_mean` double unsigned NOT NULL,

`band10\_var` double unsigned NOT NULL,

`frequency\_mean` double unsigned NOT NULL,

`frequency\_var` double unsigned NOT NULL,

`ec\_mean` double unsigned NOT NULL,

`ec\_var` double unsigned NOT NULL,

`tnp\_mean` double unsigned NOT NULL,

`tnp\_var` double unsigned NOT NULL,

`edsp\_mean` double unsigned NOT NULL,

`edsp\_var` double unsigned NOT NULL,

`fdsp\_mean` double unsigned NOT NULL,

`fdsp\_var` double unsigned NOT NULL,

`edsnr\_mean` double unsigned NOT NULL,

`edsnr\_var` double unsigned NOT NULL,

`fdsnr\_mean` double unsigned NOT NULL,

`fdsnr\_var` double unsigned NOT NULL,

`isPulse` boolean NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`decision\_tree` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`branchID` text NOT NULL,

`data\_type` int(10) unsigned NOT NULL,

`data\_value` double unsigned NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`random\_forests` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`tree\_number` int(10) unsigned NOT NULL,

`branchID` text NOT NULL,

`data\_type` int(10) unsigned NOT NULL,

`data\_value` double unsigned NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`SVM\_alpha` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`estID` int(10) unsigned NOT NULL,

`alpha` double NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`SVM\_b` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`b` double NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;

CREATE TABLE IF NOT EXISTS qraat.`SVM\_gamma` (

`ID` int(10) unsigned NOT NULL AUTO\_INCREMENT,

`deploymentID` int(10) unsigned NOT NULL,

`siteID` int(10) unsigned NOT NULL,

`start\_time` decimal(16,6) NOT NULL,

`validation` int(10) unsigned NOT NULL,

`gamma` double unsigned NOT NULL,

PRIMARY KEY (`ID`)

) ENGINE=MyISAM;