



























remaining: 141ms 900: learn: 0.2937496 total: 1.28s remaining: 140ms 901: learn: 0.2936473 total: 1.28s 902: learn: 0.2935383 total: 1.29s remaining: 138ms 903: learn: 0.2932741 total: 1.29s remaining: 137ms 904: learn: 0.2931433 total: 1.29s remaining: 135ms 905: learn: 0.2931072 total: 1.29s remaining: 134ms 906: learn: 0.2930124 total: 1.29s remaining: 133ms 907: learn: 0.2929507 total: 1.29s remaining: 131ms total: 1.3s 908: learn: 0.2928749 remaining: 130ms 909: learn: 0.2927964 total: 1.3s remaining: 128ms 910: learn: 0.2927355 total: 1.3s remaining: 127ms total: 1.3s 911: learn: 0.2925072 remaining: 126ms 912: learn: 0.2924350 total: 1.3s remaining: 124ms 913: learn: 0.2923893 total: 1.3s remaining: 123ms 914: learn: 0.2922643 total: 1.31s remaining: 121ms 915: learn: 0.2920757 total: 1.31s remaining: 120ms 916: learn: 0.2919972 total: 1.31s remaining: 119ms 917: learn: 0.2919382 total: 1.31s remaining: 117ms 918: learn: 0.2919158 total: 1.31s remaining: 116ms 919: learn: 0.2917961 total: 1.31s remaining: 114ms 920: learn: 0.2916408 total: 1.31s remaining: 113ms 921: learn: 0.2915810 total: 1.31s remaining: 111ms 922: learn: 0.2915205 total: 1.32s remaining: 110ms 923: learn: 0.2914487 total: 1.32s remaining: 108ms 924: learn: 0.2914125 total: 1.32s remaining: 107ms total: 1.32s 925: learn: 0.2911329 remaining: 105ms total: 1.32s 926: learn: 0.2910400 remaining: 104ms 927: learn: 0.2908132 total: 1.32s remaining: 103ms remaining: 101ms 928: learn: 0.2906772 total: 1.32s 929: learn: 0.2905811 total: 1.32s remaining: 99.7ms remaining: 98.3ms 930: learn: 0.2905089 total: 1.33s remaining: 96.8ms 931: learn: 0.2903460 total: 1.33s 932: learn: 0.2901279 total: 1.33s remaining: 95.4ms total: 1.33s remaining: 94ms 933: learn: 0.2899957 total: 1.33s 934: learn: 0.2899071 remaining: 92.5ms total: 1.33s 935: learn: 0.2898269 remaining: 91.1ms total: 1.33s 936: learn: 0.2895969 remaining: 89.7ms 937: learn: 0.2894449 total: 1.33s remaining: 88.2ms 938: learn: 0.2893186 total: 1.33s remaining: 86.8ms 939: learn: 0.2892549 total: 1.34s remaining: 85.3ms 940: learn: 0.2891595 total: 1.34s remaining: 83.9ms 941: learn: 0.2891209 total: 1.34s remaining: 82.5ms 942: learn: 0.2890585 total: 1.34s remaining: 81.1ms 943: learn: 0.2890267 total: 1.34s remaining: 79.7ms 944: learn: 0.2889570 total: 1.34s remaining: 78.2ms 945: learn: 0.2888953 total: 1.34s remaining: 76.8ms 946: learn: 0.2886792 total: 1.35s remaining: 75.4ms 947: learn: 0.2884946 total: 1.35s remaining: 74ms 948: learn: 0.2883550 total: 1.35s remaining: 72.5ms 949: learn: 0.2881913 total: 1.35s remaining: 71.1ms 950: learn: 0.2880346 total: 1.35s remaining: 69.7ms 951: learn: 0.2880120 total: 1.35s remaining: 68.2ms 952: learn: 0.2878553 total: 1.35s remaining: 66.8ms 953: learn: 0.2877767 total: 1.35s remaining: 65.4ms total: 1.36s 954: learn: 0.2877628 remaining: 63.9ms 955: learn: 0.2876516 total: 1.36s remaining: 62.5ms 956: learn: 0.2875138 total: 1.36s remaining: 61.1ms 957: learn: 0.2874188 total: 1.36s remaining: 59.6ms 958: learn: 0.2873819 total: 1.36s remaining: 58.2ms 959: learn: 0.2872645 total: 1.36s remaining: 56.8ms remaining: 55.4ms 960: learn: 0.2872019 total: 1.36s learn: 0.2871630 961: total: 1.36s remaining: 53.9ms 962: learn: 0.2870670 total: 1.37s remaining: 52.5ms 963: learn: 0.2869972 total: 1.37s remaining: 51.1ms 964: learn: 0.2868638 total: 1.37s remaining: 49.6ms remaining: 48.2ms learn: 0.2866161 total: 1.37s 966: learn: 0.2865765 total: 1.37s remaining: 46.8ms remaining: 45.4ms 967: learn: 0.2863801 total: 1.37s remaining: 44ms learn: 0.2862693 total: 1.37s 968: total: 1.38s remaining: 42.5ms learn: 0.2862289 969: total: 1.38s remaining: 41.1ms learn: 0.2861831 970: remaining: 39.7ms learn: 0.2861505 971: total: 1.38s learn: 0.2861193 remaining: 38.3ms 972: total: 1.38s remaining: 36.8ms learn: 0.2859927 973: total: 1.38s learn: 0.2859507 remaining: 35.4ms 974: total: 1.38s learn: 0.2858821 remaining: 34ms 975: total: 1.38s learn: 0.2857509 remaining: 32.6ms 976: total: 1.38s learn: 0.2856414 remaining: 31.1ms 977: total: 1.38s remaining: 29.7ms learn: 0.2855376 978: total: 1.39s remaining: 28.3ms learn: 0.2854827 979: total: 1.39s learn: 0.2853861 remaining: 26.9ms 980: total: 1.39s learn: 0.2853476 total: 1.39s 981: remaining: 25.5ms learn: 0.2853124 982: total: 1.39s remaining: 24.1ms learn: 0.2852616 remaining: 22.7ms total: 1.39s 983: remaining: 21.2ms learn: 0.2852023 total: 1.4s 984: remaining: 19.8ms learn: 0.2851014 985: total: 1.4s total: 1.4s remaining: 19.8ms
total: 1.4s remaining: 18.4ms
total: 1.4s remaining: 17ms
total: 1.4s remaining: 15.6ms
total: 1.4s remaining: 14.2ms
total: 1.4s remaining: 12.7ms
total: 1.4s remaining: 11.3ms learn: 0.2850422 986: learn: 0.2849396 987: 988: learn: 0.2848950 989: learn: 0.2848518 learn: 0.2846491 990: 991: learn: 0.2845500 total: 1.41s remaining: 9.91ms learn: 0.2844121 992: total: 1.41s remaining: 8.5ms 993: learn: 0.2843443 learn: 0.2843051 total: 1.41s remaining: 7.08ms 994: total: 1.41s remaining: 5.67ms 995: learn: 0.2842434 total: 1.41s remaining: 4.25ms learn: 0.2841570 996: remaining: 2.83ms 997: learn: 0.2840381 total: 1.41s total: 1.42s remaining: 1.42ms 998: learn: 0.2839733 learn: 0.2839293 total: 1.42s remaining: Ous 999: Accuracy: 83.17 % Standard Deviation: 4.70 % Training various models on the Training set 1. Logistic Regression In [36]: from sklearn.linear\_model import LogisticRegression clf\_lr = LogisticRegression(random\_state=0) clf\_lr.fit(X\_train, y\_train) LogisticRegression(random\_state=0) In [37]: y\_pred\_lr = clf\_lr.predict(X test) 2. K-Nearest Neighbor (K-NN) In [38]: **from** sklearn.neighbors **import** KNeighborsClassifier clf knn = KNeighborsClassifier(n neighbors = 5, metric = 'minkowski', p = 2) clf knn.fit(X train, y train) Out[38]: KNeighborsClassifier() In [39]: y pred knn = clf knn.predict(X test) 3. Support Vector Machine (SVM) In [40]: **from** sklearn.svm **import** SVC clf svc = SVC(kernel='linear', random state=0) clf\_svc.fit(X\_train, y\_train) Out[40]: SVC(kernel='linear', random\_state=0) In [41]: y pred svc = clf svc.predict(X test) 4. Kernel SVM In [42]: from sklearn.svm import SVC clf\_kernelSVC = SVC(kernel='rbf', random\_state=0) clf kernelSVC.fit(X\_train, y\_train) SVC(random state=0) Out[42]: In [43]: y\_pred\_kernelSVC = clf\_kernelSVC.predict(X\_test) 5. Naïve Bayes In [44]: from sklearn.naive bayes import GaussianNB clf nb = GaussianNB() clf nb.fit(X train, y train) GaussianNB() Out[44]: In [45]: y\_pred\_nb = clf\_nb.predict(X test) 6. Decision Tree 6.1 with **GINI** In [46]: **from** sklearn.tree **import** DecisionTreeClassifier clf dtGINI = DecisionTreeClassifier(criterion='gini', random state=0) clf\_dtGINI.fit(X\_train, y\_train) DecisionTreeClassifier(random state=0) Out[46]: In [47]: y\_pred\_dtGINI = clf\_dtGINI.predict(X\_test) 6.2 with **ENTROPY** In [48]: from sklearn.tree import DecisionTreeClassifier clf dtENTROPY = DecisionTreeClassifier(criterion='entropy', random state=0) clf\_dtENTROPY.fit(X\_train, y\_train) DecisionTreeClassifier(criterion='entropy', random\_state=0) Out[48]: In [49]: y\_pred\_dtENTROPY = clf\_dtENTROPY.predict(X test) 7. Random Forest Classifier 7.2 with ENTROPY In [50]: from sklearn.ensemble import RandomForestClassifier clf\_rfcGINI = RandomForestClassifier(n\_estimators = 10, criterion = 'gini', random\_state = 0) clf\_rfcGINI.fit(X\_train, y\_train) RandomForestClassifier(n\_estimators=10, random\_state=0) Out[50]: In [51]: y\_pred\_rfcGINI = clf\_rfcGINI.predict(X\_test) 7.2 with **ENTROPY** In [52]: **from** sklearn.ensemble **import** RandomForestClassifier clf rfcENTROPY = RandomForestClassifier(n estimators = 10, criterion = 'entropy', random state = 0) clf\_rfcENTROPY.fit(X\_train, y\_train) Out[52]: RandomForestClassifier(criterion='entropy', n\_estimators=10, random\_state=0) In [53]: y\_pred\_rfcENTROPY = clf\_rfcENTROPY.predict(X test) **Evaluating the model performance with Confusion Matrix** In [54]: pip install -U prettytable Requirement already satisfied: prettytable in /opt/conda/lib/python3.7/site-packages (2.4.0) Collecting prettytable Downloading prettytable-3.0.0-py3-none-any.whl (24 kB) Requirement already satisfied: wcwidth in /opt/conda/lib/python3.7/site-packages (from prettytable) (0.2.5) Requirement already satisfied: importlib-metadata in /opt/conda/lib/python3.7/site-packages (from prettytable)  $Requirement \ already \ satisfied: \ zipp>=0.5 \ in \ /opt/conda/lib/python 3.7/site-packages \ (from \ importlib-metadata->properties of the packages \ (from \ importlib-metadata->properties \ (from \$ ettytable) (3.6.0) Requirement already satisfied: typing-extensions>=3.6.4 in /opt/conda/lib/python3.7/site-packages (from importl ib-metadata->prettytable) (3.10.0.2) Installing collected packages: prettytable Attempting uninstall: prettytable Found existing installation: prettytable 2.4.0 Uninstalling prettytable-2.4.0: Successfully uninstalled prettytable-2.4.0 Successfully installed prettytable-3.0.0 WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the sys tem package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv Note: you may need to restart the kernel to use updated packages. In [55]: **from** prettytable **import** PrettyTable from sklearn.metrics import confusion matrix, accuracy score In [56]: evaluataionTable = PrettyTable() evaluataionTable.field names = ["Model", "Confusion Matrix", "Accuracy"] evaluataionTable.add\_row(["Logistic Regression", confusion\_matrix(y\_test, y\_pred\_lr), accuracy\_score(y\_test, y\_ evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["K Nearest Neighbor", confusion\_matrix(y\_test, y\_pred\_knn), accuracy\_score(y\_test, y\_ evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["Support Vector Machine", confusion\_matrix(y\_test, y\_pred\_svc), accuracy\_score(y\_test evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["SVM Kernel", confusion\_matrix(y\_test, y\_pred\_kernelSVC), accuracy\_score(y\_test, y\_pred\_kernelSVC) evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["Naïve Bayes", confusion\_matrix(y\_test, y\_pred\_nb), accuracy\_score(y\_test, y\_pred\_nb) evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["Decision Tree (with GINI)", confusion\_matrix(y\_test, y\_pred\_dtGINI), accuracy\_score( evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["Decision Tree (with Entropy)", confusion\_matrix(y\_test, y\_pred\_dtENTROPY), accuracy\_ evaluataionTable.add row(["-----", "-----", "-----"]) evaluataionTable.add\_row(["Random Forest (with GINI)", confusion\_matrix(y\_test, y\_pred\_rfcGINI), accuracy\_score evaluataionTable.add row(["-----", "-----", "-----", "-----"]) evaluataionTable.add row(["Random Forest (with ENTROPY)", confusion matrix(y test, y pred rfcENTROPY), accuracy print(evaluataionTable) Model | Confusion Matrix | Accuracy +-----\_\_\_\_\_ | \_\_\_\_ \_\_\_\_\_ | \_\_\_\_ Support Vector Machine | [[266 0] | 1.0 | 1.0 ----------SVM Kernel | [[256 10] | 0.9019138755980861 | [ 31 121]] | -----Naïve Bayes | [[223 43] | 0.868421052631579 | [12 140]] | ----- | ------Decision Tree (with GINI) | [[218 48] | 0.7990430622009569 | [36 116]] | | Decision Tree (with Entropy) | [[221 45] | 0.8038277511961722 | [ 37 115]] | Random Forest (with GINI) | [[237 29] | 0.8444976076555024 | [ 36 116]] | | Random Forest (with ENTROPY) | [[234 32] | 0.8277511961722488 | [ 40 112]] | **Confusion Matrix:** In [57]: confusionMatrixTable = PrettyTable() confusionMatrixTable.field\_names = ["Model", "Accuracy"] confusionMatrixTable.add\_row(["Logistic Regression", confusion\_matrix(y\_test, y\_pred\_lr)]) confusionMatrixTable.add\_row(["K Nearest Neighbor", confusion\_matrix(y\_test, y\_pred\_knn)]) confusionMatrixTable.add\_row(["Support Vector Machine", confusion\_matrix(y\_test, y\_pred\_svc)]) confusionMatrixTable.add\_row(["SVM Kernel", confusion\_matrix(y\_test, y\_pred\_kernelSVC)]) confusionMatrixTable.add\_row(["Naïve Bayes", confusion\_matrix(y\_test, y\_pred\_nb)]) confusionMatrixTable.add\_row(["Decision Tree (with GINI)", confusion\_matrix(y\_test, y\_pred\_dtGINI)]) confusionMatrixTable.add\_row(["Decision Tree (with Entropy)", confusion\_matrix(y\_test, y\_pred\_dtENTROPY)]) confusionMatrixTable.add row(["Random Forest (with GINI)", confusion matrix(y test, y pred rfcGINI)]) confusionMatrixTable.add\_row(["Random Forest (with ENTROPY)", confusion\_matrix(y\_test, y\_pred\_rfcENTROPY)]) print(confusionMatrixTable) | Accuracy | ----+ Logistic Regression | [[253 13] | | [ 10 142]] | K Nearest Neighbor | [[231 35] | | [ 24 128]] | Support Vector Machine | [[266 0] | | [ 0 152]] | | [[256 10] | SVM Kernel | [ 31 121]] | | [[223 43] | Naïve Bayes | [ 12 140]] | Decision Tree (with GINI) | [[218 48] | | [ 36 116]] | Decision Tree (with Entropy) | [[221 45] | | [ 37 115]] | Random Forest (with GINI) | [[237 29] | | [ 36 116]] | Random Forest (with ENTROPY) | [[234 32] | | [ 40 112]] | **Accuracy Table:** In [58]: AccuracyTable = PrettyTable() AccuracyTable.field\_names = ["Model", "Accuracy"] AccuracyTable.add\_row(["Logistic Regression", accuracy\_score(y\_test, y\_pred\_lr)])
AccuracyTable.add\_row(["K Nearest Neighbor", accuracy\_score(y\_test, y\_pred\_knn)]) AccuracyTable.add\_row(["Support Vector Machine", accuracy\_score(y\_test, y\_pred\_svc)]) AccuracyTable.add\_row(["SVM Kernel", accuracy\_score(y\_test, y\_pred\_kernelSVC)]) AccuracyTable.add\_row(["Naïve Bayes", accuracy\_score(y\_test, y\_pred\_nb)]) AccuracyTable.add\_row(["Decision Tree (with GINI)", accuracy\_score(y\_test, y\_pred\_dtGINI)]) AccuracyTable.add\_row(["Decision Tree (with Entropy)", accuracy\_score(y\_test, y\_pred\_dtENTROPY)]) AccuracyTable.add\_row(["Random Forest (with GINI)", accuracy\_score(y\_test, y\_pred\_rfcGINI)]) AccuracyTable.add\_row(["Random Forest (with ENTROPY)", accuracy\_score(y\_test, y\_pred\_rfcENTROPY)]) print(AccuracyTable) Logistic Regression | 0.9449760765550239 | K Nearest Neighbor | 0.8588516746411483 | Support Vector Machine | 1.0 SVM Kernel | 0.9019138755980861 | Naïve Bayes | 0.868421052631579 | Decision Tree (with GINI) | 0.7990430622009569 | Decision Tree (with Entropy) | 0.8038277511961722 | | Random Forest (with GINI) | 0.8444976076555024 | | Random Forest (with ENTROPY) | 0.8277511961722488 | In [59]; output = pd.DataFrame({"PassengerId": test dataset.PassengerId, "Survived": y pred kernelSVC}) output.to csv('19BCE245 DL Prac1 kernelSVC.csv', index=False) PassengerId Survived 0 892 0 1 893 2 0 894 895 3 0 896 4 1305 0 413 1306 414 1307 415 0 1308 416 1309 417 [418 rows x 2 columns] Grid Search with SVM taking too much load. IGNORE THIS SECTION. In [60]: # from sklearn.svm import SVC # classifier = SVC(kernel = 'rbf', random state = 0) # classifier.fit(X\_train, y\_train) # from sklearn.metrics import confusion matrix, accuracy score # y pred = classifier.predict(X test) # cm = confusion\_matrix(y\_test, y\_pred) # print(cm) # accuracy\_score(y\_test, y\_pred) In [61]: # # k-fold # from sklearn.model selection import cross val score # accuracies = cross val score(estimator=classifier, X=X train, y=y train, cv=10) # print(f"Accuracy : {accuracies.mean()\*100:.2f}%") # print(f"Standard Deviation : {accuracies.std()\*100:.2f}%") In [62]: | # #grid search # from sklearn.model selection import GridSearchCV # parameters = [ 'C': np.linspace(0.01, 1,num=10).tolist(), # [0, 0.25, 0.5, 0.75, 1] 'kernel' : ['linear'], 'degree' : [3, 4, 5], }, 'C': np.linspace(0.01, 1,num=10).tolist(), # [0, 0.25, 0.5, 0.75, 1] 'kernel' : ['rbf','poly','sigmoid'], 'gamma' : np.linspace(0.01, 1,num=10).tolist(), # [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0 'degree' : [3,4,5], # grid search = GridSearchCV(estimator=classifier, param grid=parameters, scoring='accuracy', # As we are doing classification, we are using accuracy for s cv=10, # number of folds (same like k-fold cross validation) n jobs =- 1 # all your processors will be used available in hardware # grid search.fit(X train, y train) In [63]: # #grid search 2 (smaller one) # from sklearn.model selection import GridSearchCV # parameters = | 'C': [0, 0.25, 0.5, 0.75, 1], 'kernel' : ['linear'], # 'degree' : [3, 4, 5], # }, 'C' : [0, 0.3, 0.65, 1], 'kernel' : ['rbf','poly','sigmoid'], 'gamma': [0.1, 0.5, 0.9], 'degree' : [3, 4, 5], # grid search = GridSearchCV(estimator=classifier, param grid=parameters, scoring='accuracy', # As we are doing classification, we are using accuracy for s cv=4, # number of folds (same like k-fold cross validation) n jobs =- 1 # all your processors will be used available in hardware # grid search.fit(X\_train, y\_train) In [64]: # best accuracy = grid search.best score # print(f"Best Accuracy achieved : {best accuracy\*100:.2f}%") In [65]: # best\_parameters = grid search.best params # print(f"Best parameters achieved : {best parameters}")

890:

891:

892:

893:

894:

895:

896:

897:

898:

899:

learn: 0.2950118

learn: 0.2948808

learn: 0.2946069

learn: 0.2944556

learn: 0.2943538

learn: 0.2942827

learn: 0.2942218

learn: 0.2941523

learn: 0.2940716

learn: 0.2938460

total: 1.27s

total: 1.27s

total: 1.27s

total: 1.27s

total: 1.28s

total: 1.28s

total: 1.28s

total: 1.28s

total: 1.28s

total: 1.28s

remaining: 155ms

remaining: 154ms

remaining: 153ms

remaining: 151ms

remaining: 150ms

remaining: 148ms

remaining: 147ms

remaining: 145ms

remaining: 144ms remaining: 142ms