Machine Learning Assignment 1 Analysis

Analysis on Variation of percentage

when cross\_validation=5, and max\_depth=1

output: array([0.19166667, 0.19444444, 0.2005571 , 0.19498607, 0.19498607])

when cross\_validation=5, and max\_depth=2

output: array([0.36111111, 0.32777778, 0.34818942, 0.33983287, 0.29526462])

when cross\_validation=5, and max\_depth=3

output array([0.49166667, 0.51388889, 0.51810585, 0.52367688, 0.46796657])

when cross\_validation=5, and max\_depth=4

output: array([0.49166667, 0.51388889, 0.51810585, 0.52367688, 0.46796657])

when cross\_validation=5, and max\_depth=5

output: array ([0.71666667, 0.66388889, 0.79108635, 0.75766017, 0.74651811])

when cross\_validation=5, and max\_depth=6

output: array([0.775 , 0.70833333, 0.82451253, 0.80501393, 0.77158774])

when cross\_validation=5, and max\_depth=7

output: array([0.76666667, 0.75833333, 0.8356546 , 0.84122563, 0.80222841])

when cross\_validation=5, and max\_depth=8

output: array([0.8 , 0.75 , 0.82451253, 0.83008357, 0.79108635])

when cross\_validation=5, and max\_depth=9

output: array([0.81388889, 0.77222222, 0.83008357, 0.82729805, 0.77437326])

when cross\_validation=5, and max\_depth=10

output: array([0.81111111, 0.775 , 0.82172702, 0.83286908, 0.77994429])

Feature\_importance values for max\_depth=7:

array([0. , 0.00176884, 0.01112014, 0.0003562 , 0.00172928,

0.02421861, 0.00147502, 0. , 0. , 0.01280532,

0.00813655, 0.00106861, 0.00297481, 0.00049065, 0.0003562 ,

0. , 0. , 0. , 0.00425251, 0.03475065,

0.02253181, 0.10286839, 0. , 0. , 0. ,

0. , 0.10240376, 0.0384585 , 0.01011621, 0.00570632,

0.02305012, 0. , 0. , 0.00776335, 0.00697695,

0.0077678 , 0.09246632, 0.015313 , 0.05023051, 0. ,

0. , 0.00300646, 0.14788587, 0.09428794, 0.00086464,

0.01288521, 0.00633522, 0. , 0. , 0. ,

0.00666257, 0.0033402 , 0.00381887, 0.0367372 , 0.07090546,

0. , 0. , 0. , 0.00376845, 0.0009813 ,

0.00453065, 0.01283351, 0. , 0. ])

Feature\_importance values for max\_depth=8:

array([0. , 0. , 0.01029052, 0. , 0.00305249,

0.02722866, 0. , 0. , 0. , 0.01374918,

0.00491287, 0.00464552, 0.00218949, 0.00244029, 0. ,

0. , 0. , 0.00436747, 0.00821207, 0.0359582 ,

0.02234387, 0.10010937, 0. , 0. , 0. ,

0. , 0.09593626, 0.03759575, 0.01056307, 0.00710568,

0.02213327, 0. , 0. , 0.00839682, 0.0010261 ,

0.01036594, 0.08785845, 0.01515261, 0.04981502, 0. ,

0. , 0.00447464, 0.14200349, 0.09147977, 0.00200734,

0.01326151, 0.00535692, 0. , 0. , 0. ,

0.00817449, 0.00612563, 0.0011552 , 0.03459185, 0.06875208,

0. , 0. , 0. , 0.01556976, 0.00407444,

0.00047113, 0.01305036, 0.00400242, 0. ])

Feature\_importance values for max\_depth=9:

array([0. , 0.00067358, 0.00849338, 0.00138788, 0.00300569,

0.02921335, 0.00167244, 0. , 0. , 0.01333051,

0.0093771 , 0.00289035, 0.00282951, 0.00355131, 0.00133035,

0. , 0. , 0.00430052, 0.00517498, 0.03395289,

0.02221101, 0.0990386 , 0. , 0. , 0. ,

0.00104809, 0.09446556, 0.03768772, 0.01098657, 0.00573214,

0.02418843, 0. , 0. , 0.00734028, 0.00398803,

0.00777533, 0.08848815, 0.01380738, 0.04665689, 0. ,

0. , 0.0032408 , 0.13982658, 0.09110458, 0.00109293,

0.01319339, 0.00715133, 0. , 0. , 0.00033679,

0.0090873 , 0.0041114 , 0.00486988, 0.03599976, 0.06764919,

0.0014885 , 0. , 0. , 0.01195676, 0.00401198,

0.00533869, 0.01247092, 0.00247126, 0. ])

Feature\_importance values for max\_depth=10:

array([0. , 0. , 0.01443387, 0. , 0.00291267,

0.02519839, 0. , 0. , 0. , 0.01559449,

0.00981884, 0.00346688, 0.00394664, 0.00172209, 0.00134155,

0. , 0. , 0.00428261, 0.00500504, 0.03393496,

0.02322335, 0.09883488, 0. , 0. , 0. ,

0.00125935, 0.09751128, 0.03786436, 0.01155314, 0.00629683,

0.02352302, 0. , 0. , 0.0117431 , 0.0039261 ,

0.01125794, 0.08658433, 0.0162869 , 0.04884704, 0. ,

0. , 0.00472943, 0.14033254, 0.08877823, 0.00552007,

0.01213227, 0.00098417, 0. , 0. , 0. ,

0.00627324, 0.00204681, 0.00665853, 0.03444349, 0.06979079,

0. , 0. , 0. , 0.00929798, 0.00067078,

0.00247462, 0.01167372, 0.00382369, 0. ])

According to my findings, if we consider digit max\_depth=7 as a predictor, it has good accuracy; however, if we consider feature importance as a parameter, max depth value 7 has better accuracy than 9 and the decision tree is normalized.