List of selected features

[1, 7, 28, 51, 76, 90, 123, 156, 161, 198, 226, 229, 231, 234, 240, 263, 288, 300, 333, 350, 374, 386, 392, 405, 443, 465, 477, 487, 495, 496, 511, 521, 549, 551, 597, 618, 631, 645, 649, 652, 659, 661, 668, 679, 681, 694, 698, 704, 718, 730, 731, 736, 742, 744, 752, 767, 774, 781, 818, 838, 865, 867, 869, 902, 949, 965, 974, 985, 1003, 1008, 1013, 1016, 1040, 1041, 1053, 1071, 1079, 1099, 1110, 1112, 1138, 1190, 1197, 1208, 1209, 1210, 1219, 1244, 1246, 1250, 1273, 1291, 1303, 1311, 1388, 1401, 1406, 1418, 1437, 1457, 1488, 1528, 1555, 1557, 1563, 1572, 1581, 1585, 1599, 1622, 1633, 1662, 1667, 1705, 1764, 1768, 1769, 1778, 1805, 1829, 1833, 1842, 1870, 1877, 1889, 1926, 1938, 1940, 1949, 1961, 1987, 1998, 2000, 2030, 2079, 2099, 2114, 2118, 2146, 2160, 2198, 2208, 2211, 2219, 2220, 2237, 2241, 2245, 2272, 2274, 2287, 2288, 2289, 2299, 2303, 2314, 2326, 2329, 2330, 2331, 2363, 2366, 2367, 2400, 2412, 2426, 2427, 2436, 2460, 2461, 2468, 2498, 2501, 2510, 2554, 2556, 2561, 2644, 2648, 2662, 2682, 2697, 2705, 2722, 2723, 2761, 2767, 2771, 2783, 2792, 2808, 2835, 2849, 2890, 2901, 2907, 2909, 2918, 2921, 2939, 2943, 2950, 2965, 2992, 3001, 3046, 3051, 3057, 3070, 3075, 3101, 3112, 3118, 3172, 3174, 3175, 3198, 3199, 3268, 3271, 3278, 3317, 3318, 3324, 3352, 3383, 3386, 3419, 3421, 3426, 3450, 3455, 3479, 3496, 3534, 3550, 3552, 3556, 3576, 3624, 3704, 3705, 3722, 3724, 3754, 3757, 3760, 3784, 3814, 3836, 3863, 3878, 3890, 3907, 3920, 3923, 3927, 3932, 3956, 3964, 3988, 3997, 4001, 4018, 4033, 4038, 4047, 4079, 4080, 4105, 4124, 4173, 4203, 4210, 4219, 4238, 4251, 4291, 4297, 4303, 4329, 4347, 4367, 4371, 4391, 4416, 4442, 4493, 4514, 4518, 4519, 4524, 4560, 4562, 4603, 4609, 4610, 4616, 4633, 4656, 4665, 4702, 4711, 4721, 4833, 4836, 4839, 4854, 4893, 4916, 4924, 4977, 4980, 4988, 4998, 5004, 5040, 5073, 5095, 5106, 5128, 5134, 5161, 5166, 5171, 5187, 5193, 5210, 5218, 5232, 5261, 5262, 5265, 5273, 5284, 5298, 5301, 5309, 5310, 5362, 5375, 5425, 5445, 5448, 5449, 5479, 5487, 5494, 5503, 5510, 5511, 5516, 5522, 5559, 5560, 5562, 5580, 5583, 5585, 5626, 5636, 5666, 5670, 5699, 5700, 5708, 5732, 5743, 5746, 5748, 5768, 5833, 5842, 5850, 5851, 5856, 5875, 5877, 5878, 5890, 5908, 5931, 5939, 6000, 6001, 6009, 6010, 6024, 6028, 6048, 6066, 6093, 6097, 6102, 6143, 6177, 6190, 6218, 6236, 6247, 6271, 6293, 6304, 6307, 6315, 6364, 6366, 6375, 6376, 6379, 6384, 6423, 6440, 6441, 6451, 6463, 6470, 6471, 6513, 6517, 6529, 6534, 6543, 6552, 6560, 6576, 6587, 6611, 6614, 6710, 6720, 6721, 6722, 6741, 6780, 6792, 6799, 6841, 6887, 6897, 6927, 6958, 6976, 6989, 7022, 7027, 7041, 7045, 7048, 7061, 7069, 7073, 7092, 7096, 7110, 7125, 7139, 7152, 7165, 7194, 7264, 7269, 7275, 7279, 7286, 7289, 7294, 7311, 7326, 7354, 7363, 7383, 7401, 7407, 7411, 7427, 7442, 7444, 7455, 7459, 7519, 7578, 7595, 7599, 7629, 7651, 7675, 7678, 7693, 7729, 7737, 7790, 7805, 7812, 7813, 7829, 7845, 7860, 7888, 7890, 7898, 7905, 7976, 7981, 8018, 8021, 8044, 8060, 8078, 8085, 8099, 8117, 8136, 8160, 8161, 8170, 8179, 8181, 8183, 8187, 8222, 8266, 8316, 8338, 8359, 8406, 8461, 8465, 8550, 8575, 8595, 8617, 8621, 8635, 8658, 8661, 8700, 8732, 8741, 8759, 8770, 8794, 8811, 8947, 8954, 8955, 8957, 8961, 8968, 8983, 8987, 8994, 9007, 9010, 9026, 9034, 9036, 9052, 9065, 9067, 9083, 9114, 9141, 9191, 9201, 9208, 9222, 9234, 9241, 9247, 9264, 9265, 9282, 9315, 9317, 9354, 9367, 9417, 9426, 9434, 9478, 9489, 9496, 9499, 9514, 9522, 9523, 9539, 9546, 9550, 9551, 9552, 9556, 9572, 9583, 9584, 9605, 9616, 9626, 9630, 9646, 9651, 9669, 9679, 9680, 9681, 9682, 9695, 9742, 9745, 9757, 9786, 9797, 9800, 9802, 9822, 9828, 9873, 9892, 9935, 9960, 10027, 10051, 10119, 10124, 10125, 10129, 10151, 10153, 10159, 10166, 10181, 10214, 10233, 10249, 10254, 10269, 10305, 10316, 10330, 10367, 10371, 10392, 10401, 10406, 10413, 10415, 10421, 10425, 10434, 10446, 10477, 10491, 10515, 10521, 10523, 10569, 10573, 10574, 10575, 10593, 10594, 10626, 10653, 10672, 10675, 10720, 10732, 10749, 10751, 10760, 10761, 10776, 10803, 10810, 10857, 10882, 10884, 10919, 10931, 10934, 10935, 10945, 10953, 10992, 11015, 11046, 11050, 11055, 11057, 11069, 11101, 11107, 11118, 11154, 11206, 11215, 11217, 11233, 11242, 11245, 11250, 11255, 11268, 11278, 11318, 11336, 11349, 11412, 11424, 11432, 11445, 11449, 11453, 11462, 11463, 11474, 11480, 11491, 11522, 11581, 11598, 11601, 11602, 11610, 11651, 11669, 11685, 11701, 11733, 11800, 11813, 11817, 11818, 11819, 11842, 11854, 11885, 11895, 11916, 11949, 11964, 11968, 12040, 12086, 12103, 12143, 12161, 12180, 12183, 12184, 12207, 12215, 12223, 12246, 12257, 12281, 12290, 12292, 12294, 12296, 12301, 12303, 12339, 12341, 12346, 12362, 12383, 12406, 12436, 12439, 12457, 12459, 12475, 12517, 12531, 12600, 12603, 12611, 12619, 12639, 12653, 12659, 12667, 12684, 12691, 12700, 12756, 12758, 12787, 12806, 12826, 12827, 12867, 12879, 12940, 12942, 12945, 12946, 12952, 12960, 12964, 12969, 13011, 13042, 13060, 13066, 13084, 13109, 13110, 13132, 13142, 13166, 13168, 13232, 13278, 13290, 13318, 13321, 13332, 13340, 13348, 13349, 13350, 13374, 13393, 13406, 13416, 13424, 13426, 13443, 13481, 13482, 13484, 13490, 13513, 13524, 13540, 13541, 13555, 13556, 13584, 13612, 13623, 13627, 13634, 13635, 13666, 13674, 13678, 13679, 13692, 13737, 13742, 13795, 13801, 13804, 13805, 13853, 13864, 13875, 13922, 13927, 14001, 14023, 14027, 14032, 14039, 14057, 14064, 14072, 14129, 14181, 14198, 14200, 14210, 14221, 14253, 14262, 14275, 14282, 14322, 14325, 14332, 14356, 14372, 14387, 14402, 14407, 14427, 14440, 14472, 14483, 14496, 14497, 14536, 14547, 14566, 14570, 14571, 14605, 14610, 14614, 14634, 14651, 14676, 14684, 14700, 14746, 14781, 14813, 14827, 14906, 14914, 14922, 14924, 14927, 14939, 14971, 14979, 15009, 15012, 15019, 15035, 15046, 15090, 15091, 15093, 15096, 15103, 15113, 15126, 15153, 15166, 15176, 15179, 15232, 15251, 15274, 15312, 15323, 15396, 15413, 15418, 15420, 15447, 15448, 15450, 15480, 15492, 15516, 15565, 15573, 15579, 15580, 15635, 15724, 15735, 15745, 15767, 15774, 15776, 15818, 15820, 15834, 15835, 15847, 15886, 15889, 15896, 15952, 15953, 15980, 15993, 16010, 16011, 16020, 16024, 16033, 16083, 16114, 16120, 16125, 16136, 16167, 16189, 16231, 16247, 16266, 16267, 16276, 16289, 16311, 16317, 16329, 16346, 16350, 16351, 16364, 16365]

List of Constant Feaute (not Needed)

[ 41 144 283 284 297 321 365 565 732 824 825 839

1077 1100 1188 1255 1256 1259 1338 1363 1367 1399 1425 1430

1461 1619 1655 1743 1766 1891 1993 1994 2003 2031 2032 2224

2225 2231 2236 2248 2252 2409 2476 2691 2693 2698 2699 2733

2889 2903 3039 3048 3049 3076 3078 3079 3273 3284 3288 3425

3463 3895 3994 4036 4050 4051 4052 4135 4933 5009 5033 5242

5299 5371 5390 5418 5439 5606 5607 5609 5645 5648 5755 6030

6049 6051 6118 6125 6147 6323 6324 6825 6852 6855 6856 7207

7218 7357 7403 7420 7421 7440 7602 7800 7857 7858 7913 8000

8025 8107 8565 9029 9200 9305 9485 9621 9625 9666 9755 10262

10315 10458 10464 10502 10527 10528 10585 10605 10608 10668 10690 10958

10959 10962 10963 10964 10965 10966 10996 10999 11124 11125 11192 11372

11776 11905 12140 12144 12231 12585 12593 12628 12680 12709 12752 12820

12822 12829 13030 13032 13108 13189 13236 13297 13303 13304 13307 13322

13472 13608 13629 13667 13668 13746 13862 13866 13882 13905 13908 13909

13911 13912 13918 13963 13989 13996 13997 14006 14029 14122 14149 14490

14607 14618 14639 14760 14765 15189 15192 15202 15205 15209 15237 15259

15304 15328 15350 15460 15481 15538 15654 15887 15916 15938 15940 16142

16160 16198 16241 16315 16324]

The **SelectKBest** method, particularly when combined with statistical scoring functions such as **f\_classif**, **chi2**, or **mutual\_info\_classif**, relies on statistical methods to evaluate the importance of features for classification tasks. Here's a brief overview of these statistical methods:

1. **f\_classif (ANOVA F-value for classification)**:
   * This method computes the ANOVA F-value between each feature and the target variable.
   * It assumes that the distributions of the features across different classes are similar, and it tests whether the means of the features are significantly different for different classes.
   * Features with higher F-values and lower p-values (indicating higher significance) are considered more important.
2. **chi2 (Chi-squared statistics)**:
   * Chi-squared statistics measure the dependence between the feature and the target variable for classification tasks.
   * It is particularly useful for categorical features and discrete target variables.
   * It quantifies the difference between the observed and expected frequencies of the feature values and the target classes.
   * Higher chi-squared values indicate higher dependence between the feature and the target variable.
3. **mutual\_info\_classif (Mutual information)**:
   * Mutual information measures the mutual dependence between two random variables, in this case, a feature and the target variable.
   * It captures both linear and non-linear relationships between the variables.
   * It is computed using entropy-based methods, quantifying the amount of information obtained about one variable through the other.
   * Features with higher mutual information values are considered more informative for classification.

These statistical methods help **SelectKBest** identify the most relevant features by evaluating their relationships with the target variable. By selecting features with higher scores, **SelectKBest** aims to improve the performance of machine learning models by focusing on the most informative features for classification tasks.

The provided code defines a convolutional neural network (CNN) model using TensorFlow and Keras for binary classification tasks. This model consists of several layers that process input data and output predictions. Here's a breakdown of the model and its usage:

1. **Model Architecture**:
   * The model is defined as a sequential stack of layers using **tf.keras.Sequential**.
   * The input layer reshapes the input data into a suitable format for the convolutional layers.
   * Two convolutional layers (**Conv1D**) with 64 and 128 filters respectively are applied, followed by max-pooling layers (**MaxPooling1D**) to downsample the features.
   * After flattening the output, three fully connected (**Dense**) layers are added with 256, 128, and 64 units respectively. The activation functions used are ReLU for the first and third layers, and hyperbolic tangent (tanh) for the second and fourth layers.
   * Finally, a single unit dense layer with a sigmoid activation function is added for binary classification, producing the output probability.
2. **Model Compilation**:
   * The model is compiled using **model.compile**, where the optimizer (Adam) and loss function (binary cross-entropy) are specified. Additionally, accuracy is chosen as the metric to monitor during training.
   * Alternatively, the commented-out line shows an example of compiling the model with a different optimizer (**adamax**) and loss function (**hinge**) with area under the curve (AUC) as an additional metric.
3. **Training and Evaluation**:
   * The compiled model is trained using the **fit** method on training data (**x\_train**, **y\_train**) for a specified number of epochs and batch size.
   * During training, validation data (**x\_test**, **y\_test**) can be optionally provided to evaluate the model's performance on unseen data.
   * After training, the model's performance is evaluated on the test data using the **evaluate** method, which returns the loss and accuracy metrics.
   * Finally, the test accuracy is printed to assess the model's effectiveness in making accurate predictions on unseen data.

After applying CNN on the most selected 1000 feature we get accuracy of 87.22

Model: "sequential"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

reshape (Reshape) (None, 1000, 1) 0

conv1d (Conv1D) (None, 998, 32) 128

max\_pooling1d (MaxPooling1 (None, 499, 32) 0

D)

conv1d\_1 (Conv1D) (None, 497, 64) 6208

max\_pooling1d\_1 (MaxPoolin (None, 248, 64) 0

g1D)

flatten (Flatten) (None, 15872) 0

dense (Dense) (None, 128) 2031744

dense\_1 (Dense) (None, 1) 129

=================================================================

Total params: 2038209 (7.78 MB)

Trainable params: 2038209 (7.78 MB)

Non-trainable params: 0 (0.00 Byte)

Accuracy = 87.04

Selected most important 10 biomarkers:

ADH4

LLGL2

SLC22A12

GLYAT

APOB

CAV2

GPAM

GLP2R

EPB42

HSD17B13

A diagram of a computer

Description automatically generated