**Data preprocessing on dataset of Diabetes**

**Team Members:**

1. Safia Shabbir (P19101062)
2. Erum Saleem (P19101015)
3. Hina Jameel (P19101022)
4. Hina Saeed (P19101023)

**Masters in computer science**

**2019/2020**

**UBIT**

**University of Karachi**

# **Diabetes Dataset**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pregnancies | Glucose | BloodPressure | SkinThickness | Insulin | BMI | DiabetesPedigreeFunction | Age | Outcome |
| 6 |  | 72 |  | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 1 |  | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 3 | 78 | 50 |  | 88 | 31 | 0.248 | 26 | 1 |
| 10 |  | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | 1 |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| 10 |  | 74 | 0 | 0 | 38 | 0.537 | 34 | 1 |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | 0 |
| 1 |  | 60 |  | 846 | 30.1 | 0.398 | 59 | 1 |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | 1 |
| 7 |  | 0 | 0 | 0 | 30 | 0.484 | 32 | 1 |
| 0 | 118 | 84 | 47 | 230 | 45.8 | 0.551 | 31 | 1 |
| 7 | 107 | 74 | 0 | 0 | 29.6 | 0.254 | 31 | 1 |
| 1 |  | 30 |  | 83 | 43.3 | 0.183 | 33 | 0 |
| 1 | 115 | 70 | 30 | 96 | 34.6 | 0.529 | 32 | 1 |
| 3 |  | 88 |  | 235 | 39.3 | 0.704 | 27 | 0 |
| 8 | 99 | 84 | 0 | 0 | 35.4 | 0.388 | 50 | 0 |
| 7 | 196 | 90 |  | 0 | 39.8 | 0.451 | 41 | 1 |
| 9 |  | 80 | 35 | 0 | 29 | 0.263 | 29 | 1 |
| 11 | 143 | 94 | 33 | 146 | 36.6 | 0.254 | 51 | 1 |
| 10 | 125 | 70 | 26 | 115 | 31.1 | 0.205 | 41 | 1 |
| 7 |  | 76 |  | 0 | 39.4 | 0.257 | 43 | 1 |
| 1 | 97 | 66 | 15 | 140 | 23.2 | 0.487 | 22 | 0 |
| 13 |  | 82 | 19 | 110 | 22.2 | 0.245 | 57 | 0 |
| 5 | 117 | 92 | 0 | 0 | 34.1 | 0.337 | 38 | 0 |
| 5 | 109 | 75 | 26 | 0 | 36 | 0.546 | 60 | 0 |
| 3 |  | 76 |  | 245 | 31.6 | 0.851 | 28 | 1 |
| 3 | 88 | 58 |  | 54 | 24.8 | 0.267 | 22 | 0 |
| 6 |  | 92 | 0 | 0 | 19.9 | 0.188 | 28 | 0 |
| 10 | 122 | 78 | 31 | 0 | 27.6 | 0.512 | 45 | 0 |
| 4 |  | 60 | 33 | 192 | 24 | 0.966 | 33 | 0 |
| 11 | 138 | 76 |  | 0 | 33.2 | 0.42 | 35 | 0 |
| 9 |  | 76 | 37 | 0 | 32.9 | 0.665 | 46 | 1 |
| 2 | 90 | 68 |  | 0 | 38.2 | 0.503 | 27 | 1 |
| 4 | 111 | 72 | 47 | 207 | 37.1 | 1.39 | 56 | 1 |
| 3 |  | 64 | 25 | 70 | 34 | 0.271 | 26 | 0 |
| 7 | 133 | 84 | 0 | 0 | 40.2 | 0.696 | 37 | 0 |
| 7 |  | 92 | 18 | 0 | 22.7 | 0.235 | 48 | 0 |
| 9 | 171 | 110 | 24 | 240 | 45.4 | 0.721 | 54 | 1 |
| 7 |  | 64 | 0 | 0 | 27.4 | 0.294 | 40 | 0 |
| 0 | 180 | 66 |  | 0 | 42 | 1.893 | 25 | 1 |
| 1 | 146 | 56 | 0 | 0 | 29.7 | 0.564 | 29 | 0 |
| 2 |  | 70 | 27 | 0 | 28 | 0.586 | 22 | 0 |
| 7 | 103 | 66 |  | 0 | 39.1 | 0.344 | 31 | 1 |
| 7 | 105 | 0 | 0 | 0 | 0 | 0.305 | 24 | 0 |
| 1 |  | 80 | 11 | 82 | 19.4 | 0.491 | 22 | 0 |
| 1 |  | 50 | 15 | 36 | 24.2 | 0.526 | 26 | 0 |
| 5 | 88 | 66 | 21 | 23 | 24.4 | 0.342 | 30 | 0 |
| 8 |  | 90 | 34 | 300 | 33.7 | 0.467 | 58 | 1 |
| 7 | 150 | 66 | 42 | 342 | 34.7 | 0.718 | 42 | 0 |
| 1 | 73 | 50 | 10 | 0 | 23 | 0.248 | 21 | 0 |
| 7 |  | 68 | 39 | 304 | 37.7 | 0.254 | 41 | 1 |
| 0 | 100 | 88 | 60 | 110 | 46.8 | 0.962 | 31 | 0 |
| 0 | 146 | 82 | 0 | 0 | 40.5 | 1.781 | 44 | 0 |
| 0 | 105 | 64 | 41 | 142 | 41.5 | 0.173 | 22 | 0 |
| 2 |  | 0 | 0 | 0 | 0 | 0.304 | 21 | 0 |
| 8 | 133 | 72 | 0 | 0 | 32.9 | 0.27 | 39 | 1 |
| 5 | 44 | 62 |  | 0 | 25 | 0.587 | 36 | 0 |
| 2 |  | 58 | 34 | 128 | 25.4 | 0.699 | 24 | 0 |
| 7 | 114 | 66 | 0 | 0 | 32.8 | 0.258 | 42 | 1 |
| 5 |  | 74 | 27 | 0 | 29 | 0.203 | 32 | 0 |
| 0 |  | 88 | 30 | 0 | 32.5 | 0.855 | 38 | 1 |
| 2 | 109 | 92 | 0 | 0 | 42.7 | 0.845 | 54 | 0 |
| 1 | 95 | 66 | 13 | 38 | 19.6 | 0.334 | 25 | 0 |
| 4 |  | 85 | 27 | 100 | 28.9 | 0.189 | 27 | 0 |
| 2 | 100 | 66 | 20 | 90 | 32.9 | 0.867 | 28 | 1 |
| 5 | 139 | 64 | 35 | 140 | 28.6 | 0.411 | 26 | 0 |
| 13 | 126 | 90 | 0 | 0 | 43.4 | 0.583 | 42 | 1 |
| 4 |  | 86 | 20 | 270 | 35.1 | 0.231 | 23 | 0 |
| 1 | 79 | 75 |  | 0 | 32 | 0.396 | 22 | 0 |
| 1 | 0 | 48 | 20 | 0 | 24.7 | 0.14 | 22 | 0 |
| 7 | 62 | 78 | 0 | 0 | 32.6 | 0.391 | 41 | 0 |
| 5 |  | 72 |  | 0 | 37.7 | 0.37 | 27 | 0 |
| 0 | 131 | 0 | 0 | 0 | 43.2 | 0.27 | 26 | 1 |
| 2 |  | 66 | 22 | 0 | 25 | 0.307 | 24 | 0 |
| 3 | 113 | 44 | 13 | 0 | 22.4 | 0.14 | 22 | 0 |
| 2 | 74 | 0 | 0 | 0 | 0 | 0.102 | 22 | 0 |
| 7 |  | 78 | 26 | 71 | 29.3 | 0.767 | 36 | 0 |
| 0 | 101 | 65 |  | 0 | 24.6 | 0.237 | 22 | 0 |
| 5 | 137 | 108 | 0 | 0 | 48.8 | 0.227 | 37 | 1 |
| 2 |  | 74 |  | 125 | 32.4 | 0.698 | 27 | 0 |
| 13 | 106 | 72 |  | 0 | 36.6 | 0.178 | 45 | 0 |
| 2 |  | 68 | 25 | 71 | 38.5 | 0.324 | 26 | 0 |
| 15 | 136 | 70 | 32 | 110 | 37.1 | 0.153 | 43 | 1 |
| 1 |  | 68 | 19 | 0 | 26.5 | 0.165 | 24 | 0 |
| 1 | 80 | 55 | 0 | 0 | 19.1 | 0.258 | 21 | 0 |
| 4 | 123 | 80 | 15 | 176 | 32 | 0.443 | 34 | 0 |
| 7 |  | 78 | 40 | 48 | 46.7 | 0.261 | 42 | 0 |
| 4 | 134 | 72 | 0 | 0 | 23.8 | 0.277 | 60 | 1 |
| 2 |  | 82 | 18 | 64 | 24.7 | 0.761 | 21 | 0 |
| 6 |  | 72 |  | 228 | 33.9 | 0.255 | 40 | 0 |
| 2 | 92 | 62 | 28 | 0 | 31.6 | 0.13 | 24 | 0 |
| 1 | 71 | 48 |  | 76 | 20.4 | 0.323 | 22 | 0 |
| 6 | 93 | 50 | 30 | 64 | 28.7 | 0.356 | 23 | 0 |
| 1 |  | 90 | 51 | 220 | 49.7 | 0.325 | 31 | 1 |
| 1 | 163 | 72 |  | 0 | 39 | 1.222 | 33 | 1 |
| 1 | 151 | 60 |  | 0 | 26.1 | 0.179 | 22 | 0 |
| 0 |  | 96 |  | 0 | 22.5 | 0.262 | 21 | 0 |
| 1 |  | 72 |  | 40 | 26.6 | 0.283 | 24 | 0 |
| 2 | 85 | 65 | 0 | 0 | 39.6 | 0.93 | 27 | 0 |
| 1 | 126 | 56 | 29 | 152 | 28.7 | 0.801 | 21 | 0 |
| 1 | 96 | 122 | 0 | 0 | 22.4 | 0.207 | 27 | 0 |
| 4 |  | 58 | 28 | 140 | 29.5 | 0.287 | 37 | 0 |
| 3 | 83 | 58 | 31 | 18 | 34.3 | 0.336 | 25 | 0 |
| 0 | 95 | 85 |  | 36 | 37.4 | 0.247 | 24 | 1 |
| 3 |  | 72 |  | 135 | 33.3 | 0.199 | 24 | 1 |
| 8 |  | 62 | 26 | 495 | 34 | 0.543 | 46 | 1 |
| 1 | 89 | 76 | 34 | 37 | 31.2 | 0.192 | 23 | 0 |
| 4 | 76 | 62 | 0 | 0 | 34 | 0.391 | 25 | 0 |
| 7 | 160 | 54 |  | 175 | 30.5 | 0.588 | 39 | 1 |
| 4 | 146 | 92 | 0 | 0 | 31.2 | 0.539 | 61 | 1 |
| 5 |  | 74 | 0 | 0 | 34 | 0.22 | 38 | 1 |
| 5 | 78 | 48 | 0 | 0 | 33.7 | 0.654 | 25 | 0 |
| 4 | 97 | 60 | 23 | 0 | 28.2 | 0.443 | 22 | 0 |
| 4 | 99 | 76 |  | 51 | 23.2 | 0.223 | 21 | 0 |
| 0 | 162 | 76 | 56 | 100 | 53.2 | 0.759 | 25 | 1 |
| 6 |  | 64 | 39 | 0 | 34.2 | 0.26 | 24 | 0 |
| 2 |  | 74 | 30 | 100 | 33.6 | 0.404 | 23 | 0 |
| 5 | 132 | 80 |  | 0 | 26.8 | 0.186 | 69 | 0 |
| 0 | 113 | 76 | 0 | 0 | 33.3 | 0.278 | 23 | 1 |
| 1 | 88 | 30 | 42 | 99 | 55 | 0.496 | 26 | 1 |
| 3 | 120 | 70 | 30 | 135 | 42.9 | 0.452 | 30 | 0 |
| 1 | 118 | 58 |  | 94 | 33.3 | 0.261 | 23 | 0 |
| 1 |  | 88 |  | 145 | 34.5 | 0.403 | 40 | 1 |
| 0 | 105 | 84 | 0 | 0 | 27.9 | 0.741 | 62 | 1 |
| 4 | 173 | 70 |  | 168 | 29.7 | 0.361 | 33 | 1 |
| 9 | 122 | 56 | 0 | 0 | 33.3 | 1.114 | 33 | 1 |
| 3 | 170 | 64 | 37 | 225 | 34.5 | 0.356 | 30 | 1 |
| 8 | 84 | 74 | 31 | 0 | 38.3 | 0.457 | 39 | 0 |
| 2 | 96 | 68 | 13 | 49 | 21.1 | 0.647 | 26 | 0 |
| 2 | 125 | 60 |  | 140 | 33.8 | 0.088 | 31 | 0 |
| 0 |  | 70 | 26 | 50 | 30.8 | 0.597 | 21 | 0 |
| 0 | 93 | 60 | 25 | 92 | 28.7 | 0.532 | 22 | 0 |
| 0 | 129 | 80 | 0 | 0 | 31.2 | 0.703 | 29 | 0 |
| 5 |  | 72 | 29 | 325 | 36.9 | 0.159 | 28 | 0 |
| 3 | 128 | 78 | 0 | 0 | 21.1 | 0.268 | 55 | 0 |
| 5 |  | 82 | 30 | 0 | 39.5 | 0.286 | 38 | 0 |
| 2 | 108 | 52 |  | 63 | 32.5 | 0.318 | 22 | 0 |
| 10 | 108 | 66 | 0 | 0 | 32.4 | 0.272 | 42 | 1 |
| 4 | 154 | 62 |  | 284 | 32.8 | 0.237 | 23 | 0 |
| 0 | 102 | 75 | 23 | 0 | 0 | 0.572 | 21 | 0 |
| 9 | 57 | 80 | 37 | 0 | 32.8 | 0.096 | 41 | 0 |
| 2 |  | 64 |  | 119 | 30.5 | 1.4 | 34 | 0 |
| 5 | 147 | 78 | 0 | 0 | 33.7 | 0.218 | 65 | 0 |
| 2 |  | 70 | 17 | 0 | 27.3 | 0.085 | 22 | 0 |
| 1 | 136 | 74 |  | 204 | 37.4 | 0.399 | 24 | 0 |
| 4 | 114 | 65 | 0 | 0 | 21.9 | 0.432 | 37 | 0 |
| 9 |  | 86 | 28 | 155 | 34.3 | 1.189 | 42 | 1 |
| 1 | 153 | 82 |  | 485 | 40.6 | 0.687 | 23 | 0 |
| 8 |  | 78 | 0 | 0 | 47.9 | 0.137 | 43 | 1 |
| 7 | 152 | 88 |  | 0 | 50 | 0.337 | 36 | 1 |
| 2 | 99 | 52 | 15 | 94 | 24.6 | 0.637 | 21 | 0 |
| 1 |  | 56 | 21 | 135 | 25.2 | 0.833 | 23 | 0 |
| 2 | 88 | 74 |  | 53 | 29 | 0.229 | 22 | 0 |
| 17 |  | 72 | 41 | 114 | 40.9 | 0.817 | 47 | 1 |
| 4 | 151 | 90 | 38 | 0 | 29.7 | 0.294 | 36 | 0 |
| 7 | 102 | 74 | 40 | 105 | 37.2 | 0.204 | 45 | 0 |
| 0 |  | 80 | 34 | 285 | 44.2 | 0.167 | 27 | 0 |
| 2 | 100 | 64 | 23 | 0 | 29.7 | 0.368 | 21 | 0 |
| 0 | 131 | 88 | 0 | 0 | 31.6 | 0.743 | 32 | 1 |
| 6 | 104 | 74 | 18 | 156 | 29.9 | 0.722 | 41 | 1 |
| 3 | 148 | 66 | 25 | 0 | 32.5 | 0.256 | 22 | 0 |
| 4 | 120 | 68 | 0 | 0 | 29.6 | 0.709 | 34 | 0 |
| 4 | 110 | 66 | 0 | 0 | 31.9 | 0.471 | 29 | 0 |
| 3 | 111 | 90 | 12 | 78 | 28.4 | 0.495 | 29 | 0 |
| 6 | 102 | 82 | 0 | 0 | 30.8 | 0.18 | 36 | 1 |
| 6 | 134 | 70 | 23 | 130 | 35.4 | 0.542 | 29 | 1 |
| 2 | 87 | 0 | 23 | 0 | 28.9 | 0.773 | 25 | 0 |
| 1 | 79 | 60 | 42 | 48 | 43.5 | 0.678 | 23 | 0 |
| 2 | 75 | 64 | 24 | 55 | 29.7 | 0.37 | 33 | 0 |
| 8 | 179 | 72 | 42 | 130 | 32.7 | 0.719 | 36 | 1 |
| 6 | 85 | 78 | 0 | 0 | 31.2 | 0.382 | 42 | 0 |
| 0 | 129 | 110 | 46 | 130 | 67.1 | 0.319 | 26 | 1 |
| 5 | 143 | 78 | 0 | 0 | 45 | 0.19 | 47 | 0 |
| 5 | 130 | 82 | 0 | 0 | 39.1 | 0.956 | 37 | 1 |
| 6 | 87 | 80 | 0 | 0 | 23.2 | 0.084 | 32 | 0 |
| 0 | 119 | 64 | 18 | 92 | 34.9 | 0.725 | 23 | 0 |
| 1 | 0 | 74 | 20 | 23 | 27.7 | 0.299 | 21 | 0 |
| 5 | 73 | 60 | 0 | 0 | 26.8 | 0.268 | 27 | 0 |
| 4 | 141 | 74 | 0 | 0 | 27.6 | 0.244 | 40 | 0 |
| 7 | 194 | 68 | 28 | 0 | 35.9 | 0.745 | 41 | 1 |
| 8 | 181 | 68 | 36 | 495 | 30.1 | 0.615 | 60 | 1 |
| 1 | 128 | 98 | 41 | 58 | 32 | 1.321 | 33 | 1 |
| 8 | 109 | 76 | 39 | 114 | 27.9 | 0.64 | 31 | 1 |
| 5 | 139 | 80 | 35 | 160 | 31.6 | 0.361 | 25 | 1 |
| 3 | 111 | 62 | 0 | 0 | 22.6 | 0.142 | 21 | 0 |
| 9 | 123 | 70 | 44 | 94 | 33.1 | 0.374 | 40 | 0 |
| 7 | 159 | 66 | 0 | 0 | 30.4 | 0.383 | 36 | 1 |
| 11 | 135 | 0 | 0 | 0 | 52.3 | 0.578 | 40 | 1 |
| 8 | 85 | 55 | 20 | 0 | 24.4 | 0.136 | 42 | 0 |
| 5 | 158 | 84 | 41 | 210 | 39.4 | 0.395 | 29 | 1 |
| 1 | 105 | 58 | 0 | 0 | 24.3 | 0.187 | 21 | 0 |
| 3 | 107 | 62 | 13 | 48 | 22.9 | 0.678 | 23 | 1 |
| 4 | 109 | 64 | 44 | 99 | 34.8 | 0.905 | 26 | 1 |
| 4 | 148 | 60 | 27 | 318 | 30.9 | 0.15 | 29 | 1 |
| 0 | 113 | 80 | 16 | 0 | 31 | 0.874 | 21 | 0 |
| 1 | 138 | 82 | 0 | 0 | 40.1 | 0.236 | 28 | 0 |
| 0 | 108 | 68 | 20 | 0 | 27.3 | 0.787 | 32 | 0 |
| 2 | 99 | 70 | 16 | 44 | 20.4 | 0.235 | 27 | 0 |
| 6 | 103 | 72 | 32 | 190 | 37.7 | 0.324 | 55 | 0 |
| 5 | 111 | 72 | 28 | 0 | 23.9 | 0.407 | 27 | 0 |
| 8 | 196 | 76 | 29 | 280 | 37.5 | 0.605 | 57 | 1 |
| 5 | 162 | 104 | 0 | 0 | 37.7 | 0.151 | 52 | 1 |
| 1 | 96 | 64 | 27 | 87 | 33.2 | 0.289 | 21 | 0 |
| 7 | 184 | 84 | 33 | 0 | 35.5 | 0.355 | 41 | 1 |
| 2 | 81 | 60 | 22 | 0 | 27.7 | 0.29 | 25 | 0 |
| 0 | 147 | 85 | 54 | 0 | 42.8 | 0.375 | 24 | 0 |
| 7 | 179 | 95 | 31 | 0 | 34.2 | 0.164 | 60 | 0 |
| 0 | 140 | 65 | 26 | 130 | 42.6 | 0.431 | 24 | 1 |
| 9 | 112 | 82 | 32 | 175 | 34.2 | 0.26 | 36 | 1 |
| 12 | 151 | 70 | 40 | 271 | 41.8 | 0.742 | 38 | 1 |
| 5 | 109 | 62 | 41 | 129 | 35.8 | 0.514 | 25 | 1 |
| 6 | 125 | 68 | 30 | 120 | 30 | 0.464 | 32 | 0 |
| 5 | 85 | 74 | 22 | 0 | 29 | 1.224 | 32 | 1 |
| 5 | 112 | 66 | 0 | 0 | 37.8 | 0.261 | 41 | 1 |
| 0 | 177 | 60 | 29 | 478 | 34.6 | 1.072 | 21 | 1 |
| 2 | 158 | 90 | 0 | 0 | 31.6 | 0.805 | 66 | 1 |
| 7 | 119 | 0 | 0 | 0 | 25.2 | 0.209 | 37 | 0 |
| 7 | 142 | 60 | 33 | 190 | 28.8 | 0.687 | 61 | 0 |
| 1 | 100 | 66 | 15 | 56 | 23.6 | 0.666 | 26 | 0 |
| 1 | 87 | 78 | 27 | 32 | 34.6 | 0.101 | 22 | 0 |
| 0 | 101 | 76 | 0 | 0 | 35.7 | 0.198 | 26 | 0 |
| 3 | 162 | 52 | 38 | 0 | 37.2 | 0.652 | 24 | 1 |
| 4 | 197 | 70 | 39 | 744 | 36.7 | 2.329 | 31 | 0 |
| 0 | 117 | 80 | 31 | 53 | 45.2 | 0.089 | 24 | 0 |
| 4 | 142 | 86 | 0 | 0 | 44 | 0.645 | 22 | 1 |
| 6 | 134 | 80 | 37 | 370 | 46.2 | 0.238 | 46 | 1 |
| 1 | 79 | 80 | 25 | 37 | 25.4 | 0.583 | 22 | 0 |
| 4 | 122 | 68 | 0 | 0 | 35 | 0.394 | 29 | 0 |
| 3 | 74 | 68 | 28 | 45 | 29.7 | 0.293 | 23 | 0 |
| 4 | 171 | 72 | 0 | 0 | 43.6 | 0.479 | 26 | 1 |
| 7 | 181 | 84 | 21 | 192 | 35.9 | 0.586 | 51 | 1 |
| 0 | 179 | 90 | 27 | 0 | 44.1 | 0.686 | 23 | 1 |
| 9 | 164 | 84 | 21 | 0 | 30.8 | 0.831 | 32 | 1 |
| 0 | 104 | 76 | 0 | 0 | 18.4 | 0.582 | 27 | 0 |
| 1 | 91 | 64 | 24 | 0 | 29.2 | 0.192 | 21 | 0 |
| 4 | 91 | 70 | 32 | 88 | 33.1 | 0.446 | 22 | 0 |
| 3 | 139 | 54 | 0 | 0 | 25.6 | 0.402 | 22 | 1 |
| 6 | 119 | 50 | 22 | 176 | 27.1 | 1.318 | 33 | 1 |
| 2 | 146 | 76 | 35 | 194 | 38.2 | 0.329 | 29 | 0 |
| 9 | 184 | 85 | 15 | 0 | 30 | 1.213 | 49 | 1 |
| 10 | 122 | 68 | 0 | 0 | 31.2 | 0.258 | 41 | 0 |
| 0 | 165 | 90 | 33 | 680 | 52.3 | 0.427 | 23 | 0 |
| 9 | 124 | 70 | 33 | 402 | 35.4 | 0.282 | 34 | 0 |
| 1 | 111 | 86 | 19 | 0 | 30.1 | 0.143 | 23 | 0 |
| 9 | 106 | 52 | 0 | 0 | 31.2 | 0.38 | 42 | 0 |
| 2 | 129 | 84 | 0 | 0 | 28 | 0.284 | 27 | 0 |
| 2 | 90 | 80 | 14 | 55 | 24.4 | 0.249 | 24 | 0 |
| 0 | 86 | 68 | 32 | 0 | 35.8 | 0.238 | 25 | 0 |
| 12 | 92 | 62 | 7 | 258 | 27.6 | 0.926 | 44 | 1 |
| 1 | 113 | 64 | 35 | 0 | 33.6 | 0.543 | 21 | 1 |
| 3 | 111 | 56 | 39 | 0 | 30.1 | 0.557 | 30 | 0 |
| 2 | 114 | 68 | 22 | 0 | 28.7 | 0.092 | 25 | 0 |
| 1 | 193 | 50 | 16 | 375 | 25.9 | 0.655 | 24 | 0 |
| 11 | 155 | 76 | 28 | 150 | 33.3 | 1.353 | 51 | 1 |
| 3 | 191 | 68 | 15 | 130 | 30.9 | 0.299 | 34 | 0 |
| 3 | 141 | 0 | 0 | 0 | 30 | 0.761 | 27 | 1 |
| 4 | 95 | 70 | 32 | 0 | 32.1 | 0.612 | 24 | 0 |
| 3 | 142 | 80 | 15 | 0 | 32.4 | 0.2 | 63 | 0 |
| 4 | 123 | 62 | 0 | 0 | 32 | 0.226 | 35 | 1 |
| 5 | 96 | 74 | 18 | 67 | 33.6 | 0.997 | 43 | 0 |
| 0 | 138 | 0 | 0 | 0 | 36.3 | 0.933 | 25 | 1 |
| 2 | 128 | 64 | 42 | 0 | 40 | 1.101 | 24 | 0 |
| 0 | 102 | 52 | 0 | 0 | 25.1 | 0.078 | 21 | 0 |
| 2 | 146 | 0 | 0 | 0 | 27.5 | 0.24 | 28 | 1 |
| 10 | 101 | 86 | 37 | 0 | 45.6 | 1.136 | 38 | 1 |
| 2 | 108 | 62 | 32 | 56 | 25.2 | 0.128 | 21 | 0 |
| 3 | 122 | 78 | 0 | 0 | 23 | 0.254 | 40 | 0 |
| 1 | 71 | 78 | 50 | 45 | 33.2 | 0.422 | 21 | 0 |
| 13 | 106 | 70 |  | 0 | 34.2 | 0.251 | 52 | 0 |
| 2 | 100 | 70 | 52 | 57 | 40.5 | 0.677 | 25 | 0 |
| 7 | 106 | 60 | 24 | 0 | 26.5 | 0.296 | 29 | 1 |
| 0 | 104 | 64 | 23 | 116 | 27.8 | 0.454 | 23 | 0 |
| 5 | 114 | 74 | 0 | 0 | 24.9 | 0.744 | 57 | 0 |
| 2 | 108 | 62 |  | 278 | 25.3 | 0.881 | 22 | 0 |
| 0 | 146 | 70 | 0 | 0 | 37.9 | 0.334 | 28 | 1 |
| 10 | 129 | 76 | 28 | 122 | 35.9 | 0.28 | 39 | 0 |
| 7 | 133 | 88 | 15 | 155 | 32.4 | 0.262 | 37 | 0 |
| 7 | 161 | 86 |  | 0 | 30.4 | 0.165 | 47 | 1 |
| 2 | 108 | 80 | 0 | 0 | 27 | 0.259 | 52 | 1 |
| 7 | 136 | 74 |  | 135 | 26 | 0.647 | 51 | 0 |
| 5 | 155 | 84 | 44 | 545 | 38.7 | 0.619 | 34 | 0 |
| 1 | 119 | 86 | 39 | 220 | 45.6 | 0.808 | 29 | 1 |
| 4 | 96 | 56 | 17 | 49 | 20.8 | 0.34 | 26 | 0 |
| 5 | 108 | 72 | 43 | 75 | 36.1 | 0.263 | 33 | 0 |
| 0 | 78 | 88 | 29 | 40 | 36.9 | 0.434 | 21 | 0 |
| 0 | 107 | 62 | 30 | 74 | 36.6 | 0.757 | 25 | 1 |
| 2 | 128 | 78 | 37 | 182 | 43.3 | 1.224 | 31 | 1 |
| 1 | 128 | 48 | 45 | 194 | 40.5 | 0.613 | 24 | 1 |
| 0 | 161 | 50 | 0 | 0 | 21.9 | 0.254 | 65 | 0 |
| 6 | 151 | 62 | 31 | 120 | 35.5 | 0.692 | 28 | 0 |
| 2 | 146 | 70 | 38 | 360 | 28 | 0.337 | 29 | 1 |
| 0 | 126 | 84 | 29 | 215 | 30.7 | 0.52 | 24 | 0 |
| 14 | 100 | 78 | 25 | 184 | 36.6 | 0.412 | 46 | 1 |

**About this dataset**

**Context**

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes.

### Content

Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

* Pregnancies: Number of times pregnant
* Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
* BloodPressure: Diastolic blood pressure (mm Hg)
* SkinThickness: Triceps skin fold thickness (mm)
* Insulin: 2-Hour serum insulin (mu U/ml)
* BMI: Body mass index (weight in kg/(height in m)^2)
* DiabetesPedigreeFunction: Diabetes pedigree function
* Age: Age (years)
* Outcome: Class variable (0 or 1)

#### Sources:

(a) Original owners: National Institute of Diabetes and Digestive and  
Kidney Diseases  
(b) Donor of database: Vincent Sigillito (vgs@aplcen.apl.jhu.edu)  
Research Center, RMI Group Leader  
Applied Physics Laboratory  
The Johns Hopkins University  
Johns Hopkins Road  
Laurel, MD 20707  
(301) 953-6231  
(c) Date received: 9 May 1990

#### Number of Instances: 300

#### Number of Attributes: 8 plus class

#### For Each Attribute: (all numeric-valued)

1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (mu U/ml)
6. Body mass index (weight in kg/(height in m)^2)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

#### Missing Attribute Values: Yes

#### Class Distribution: (class value 1 is interpreted as "tested positive for diabetes")

#### **Coding**

library(RSQLite)

library(DBI)

library(datasets)

library(caTools)

library(e1071)

f<-file.choose("diab1.csv")

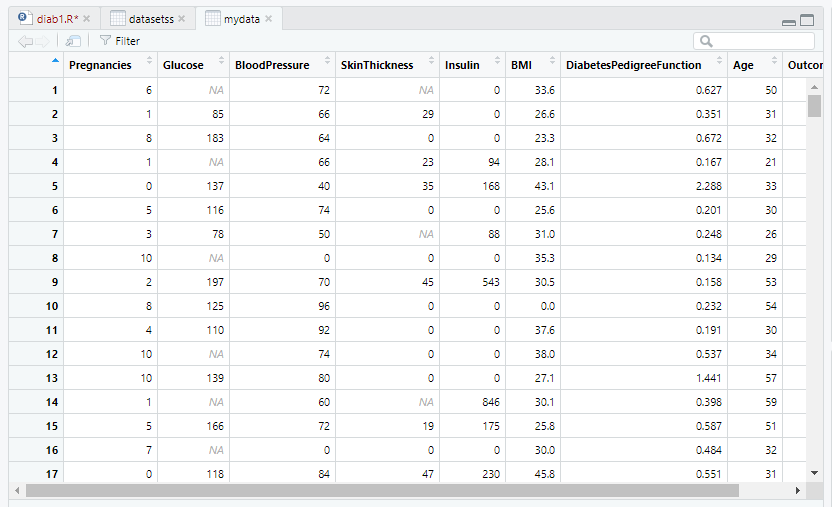
mydata<-read.csv(f)

datasetss<-read.csv(f)

#mydata<-read.csv(file = "/home/spllab01/Diabetes/diab1.csv",header = TRUE,sep = ",")

#datasetss<-read.csv(file = "/home/spllab01/Diabetes/diab1.csv",header = TRUE,sep = ",")

View(mydata)



View(datasetss)

#dataset<-read.table('diab.csv',header = T)

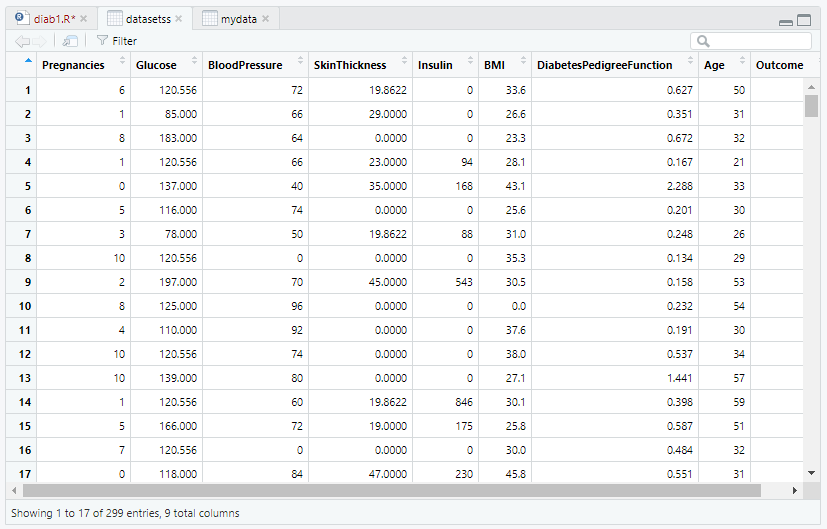
datasetss$SkinThickness = ifelse(is.na(dataset$SkinThickness), ave(dataset$SkinThickness, FUN = function(x) mean(x, na.rm = 'TRUE')), dataset$SkinThickness)

#Fills the NUll values with the average of that column values.

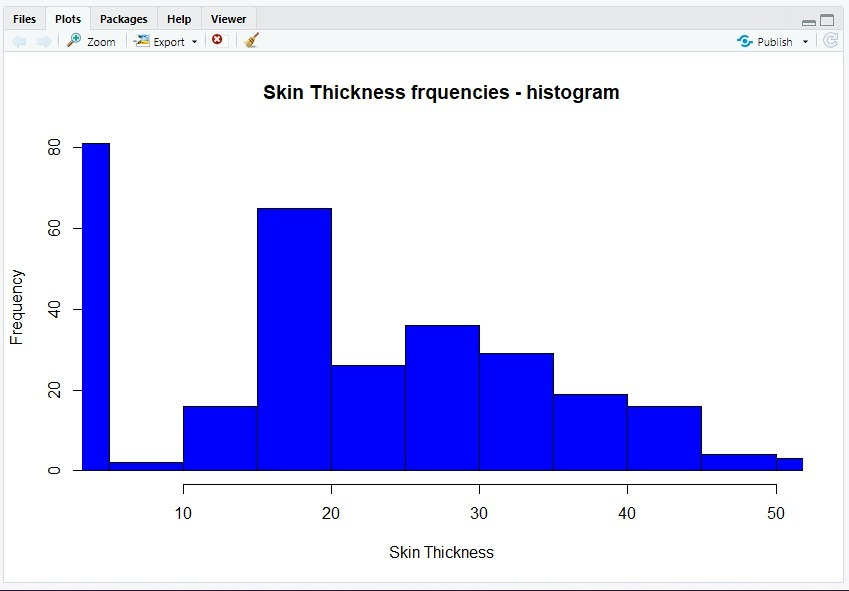
View(datasetss)

datasetss$Glucose = ifelse(is.na(dataset$Glucose), ave(dataset$Glucose, FUN = function(x) mean(x, na.rm = 'TRUE')), dataset$Glucose)

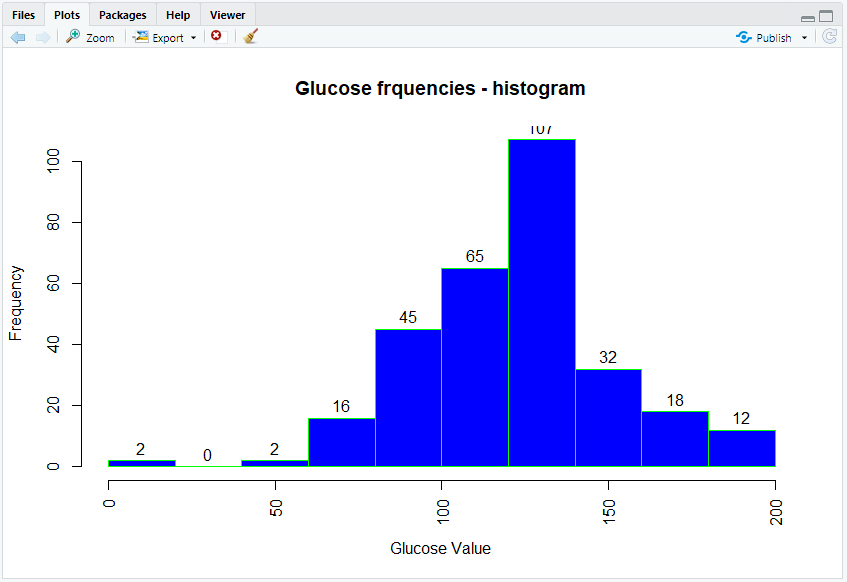
View(datasetss)



h<-hist(datasetss$SkinThickness,main="Skin Thickness frquencies - histogram", xlab = "Skin Thickness", xlim = c(5,50),col = "blue")

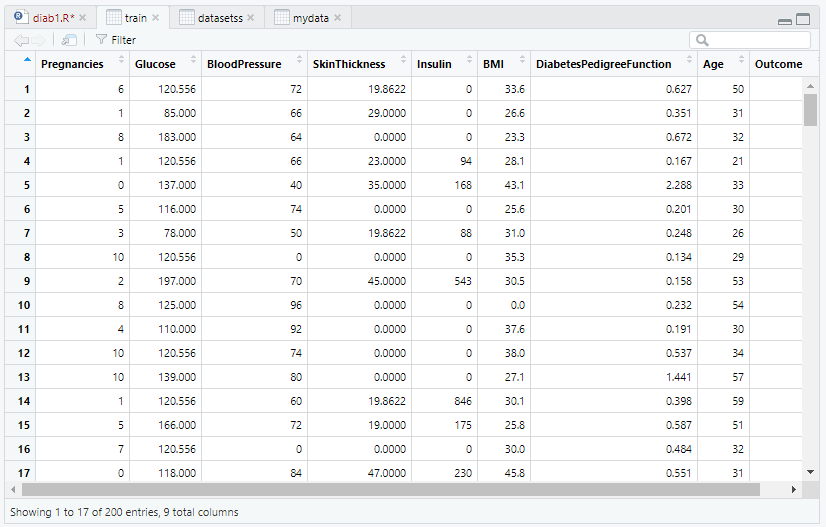


h<-hist(datasetss$Glucose,main="Glucose frquencies - histogram", xlab = "Glucose Value",col = "blue",labels = TRUE, breaks = 8, border = "green",las=3)



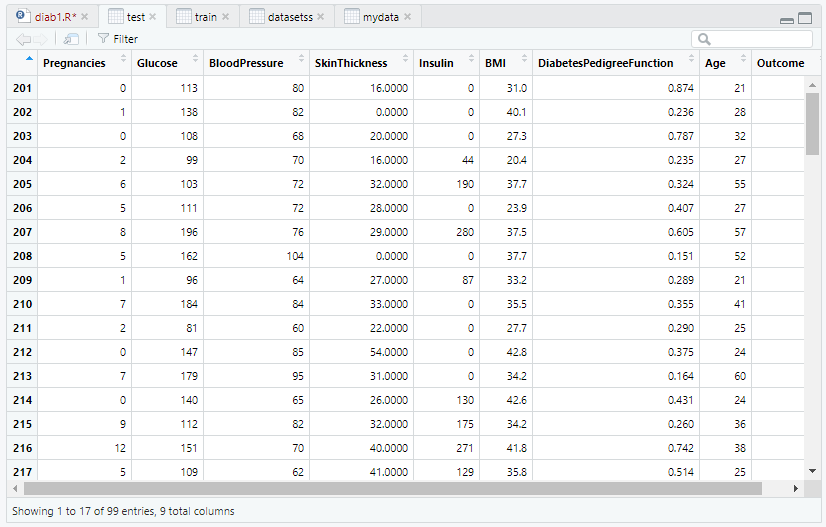
train<-as.data.frame(datasetss[1:200,])

View(train)

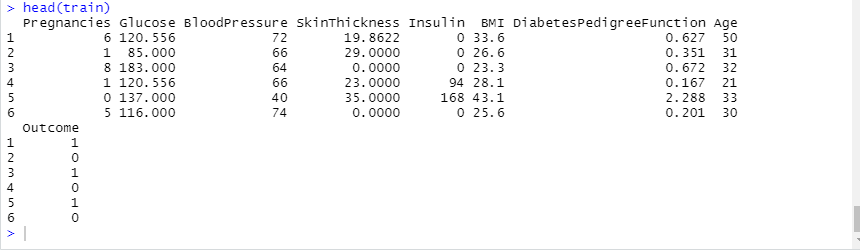


test<-as.data.frame(datasetss[201:299,])

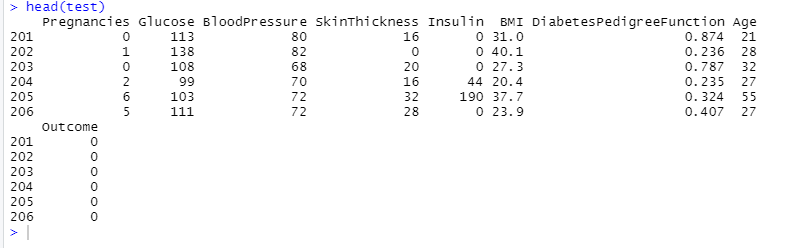
View(test)



head(train)



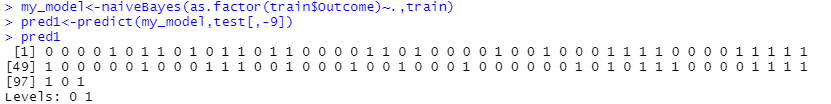
head(test)



my\_model<-naiveBayes(as.factor(train$Outcome)~.,train)

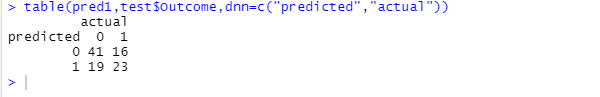
pred1<-predict(my\_model,test[,-9])

pred1



#generate the confusion matrix

table(pred1,test$Outcome,dnn=c("predicted","actual"))



#Build Classifier Models using Different Techniques.

#Cross Validation.

#Cross Validation K fold cross validation

library(caret)

library(lattice)

library(ggplot2)

# Define train control for k fold cross validation

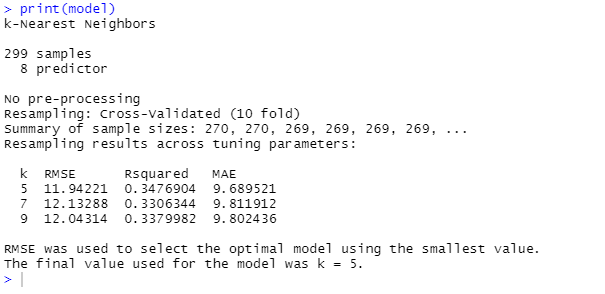
train\_control <- trainControl(method="cv", number=10)

# Fit Naive Bayes Model

model <- train(SkinThickness~., data=datasetss, trControl=train\_control, method="knn")

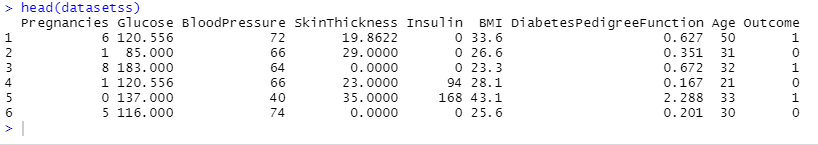
# Summarise Results

print(model)

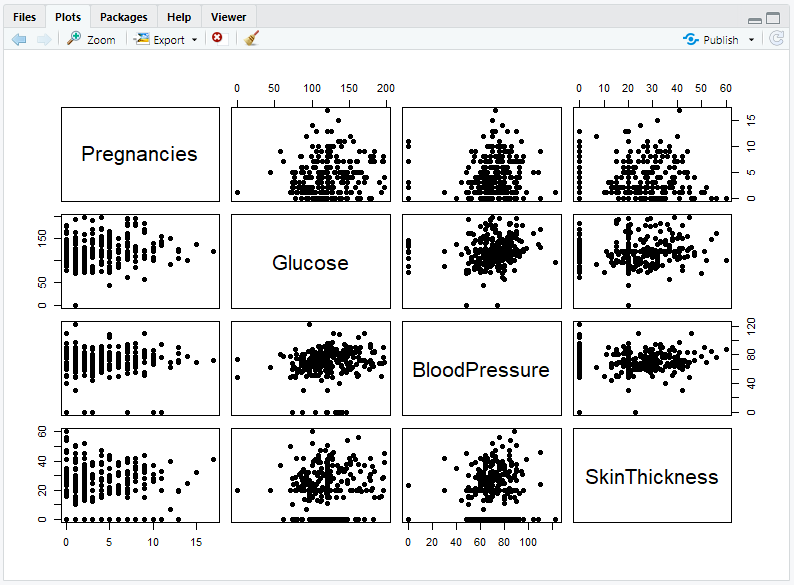


#scatterPolt matrix

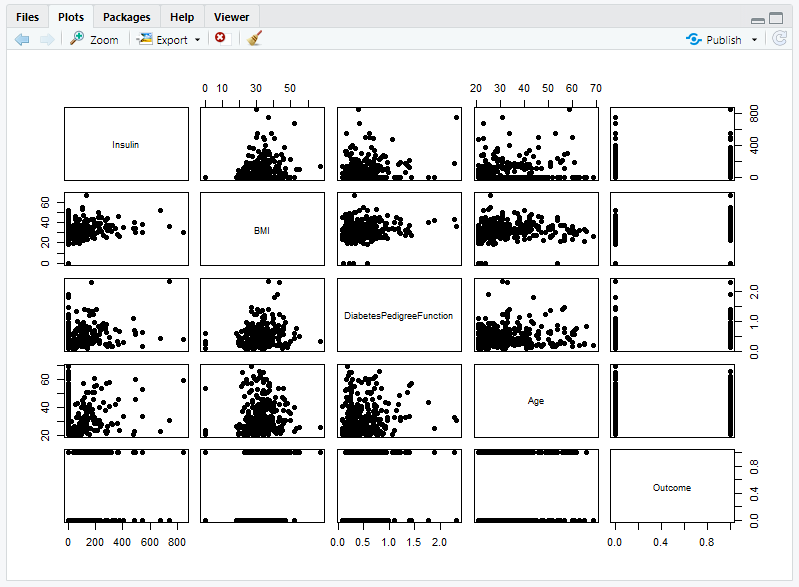
head(datasetss)



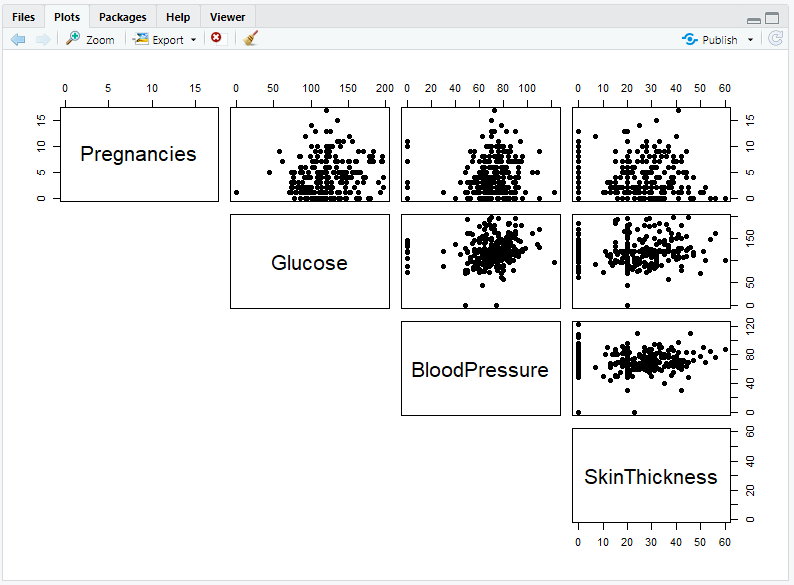
pairs(datasetss[,1:4], pch = 19)



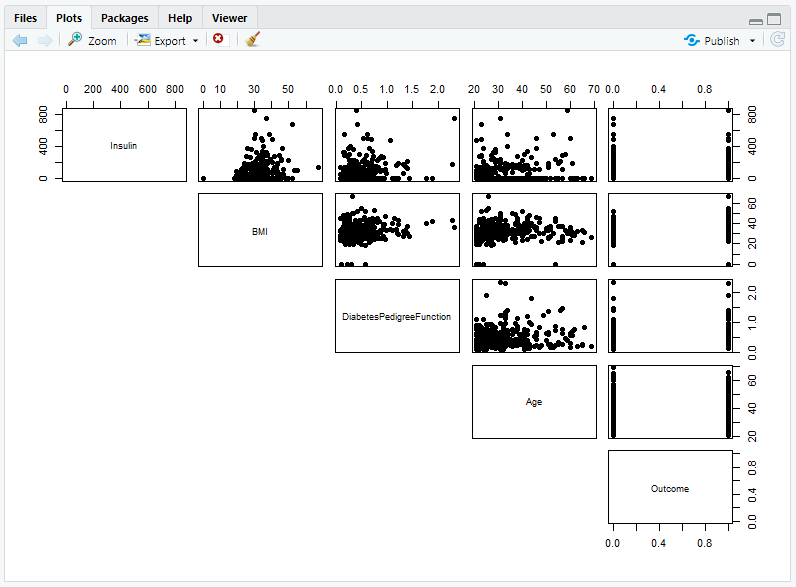
pairs(datasetss[,5:9], pch = 19)



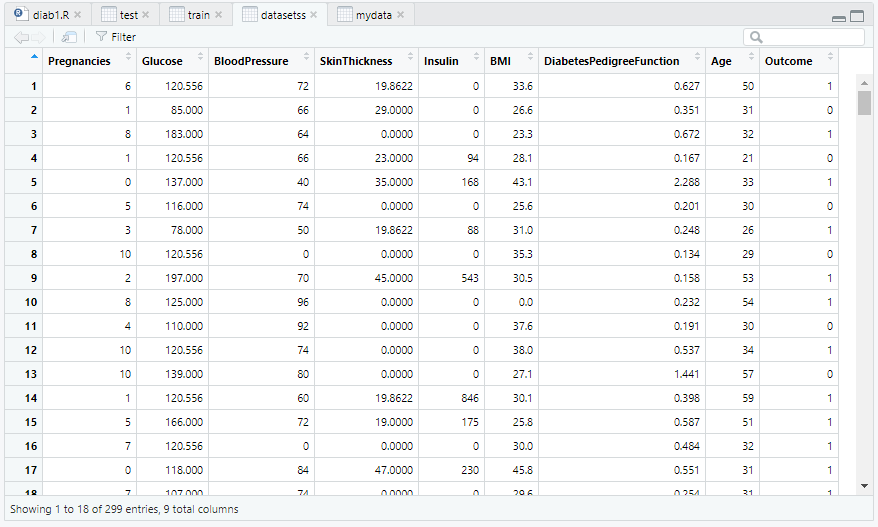
pairs(datasetss[,1:4], pch = 19, lower.panel = NULL)



pairs(datasetss[,5:9], pch = 19, lower.panel = NULL)



View(datasetss)



#One more classifier model

library(mlbench)

library(caret)

# prepare training scheme

control <- trainControl(method="repeatedcv", number=10, repeats=3)

# CART

set.seed(7)

fit.cart <- train(SkinThickness~., data=datasetss, method="rpart", trControl=control)

# SVM

set.seed(7)

fit.svm <- train(SkinThickness~., data=datasetss, method="svmRadial", trControl=control)

# kNN

set.seed(7)

fit.knn <- train(SkinThickness~., data=datasetss, method="knn", trControl=control)

# Random Forest

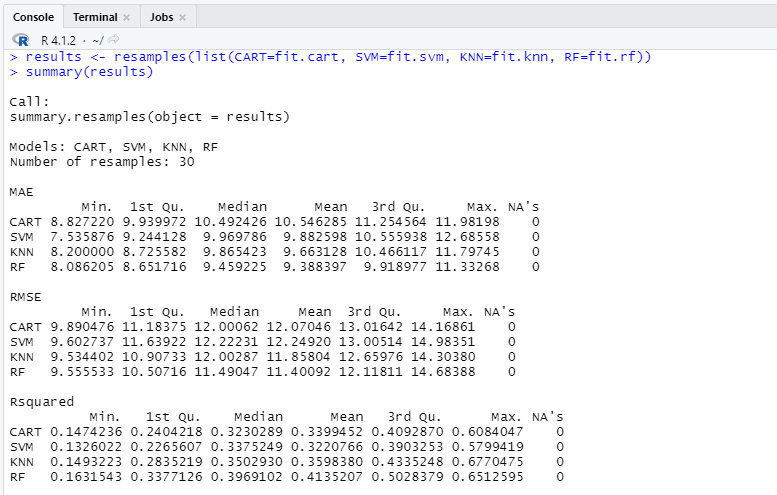
set.seed(7)

fit.rf <- train(SkinThickness~., data=datasetss, method="rf", trControl=control)

# collect resamples

results <- resamples(list(CART=fit.cart, SVM=fit.svm, KNN=fit.knn, RF=fit.rf))

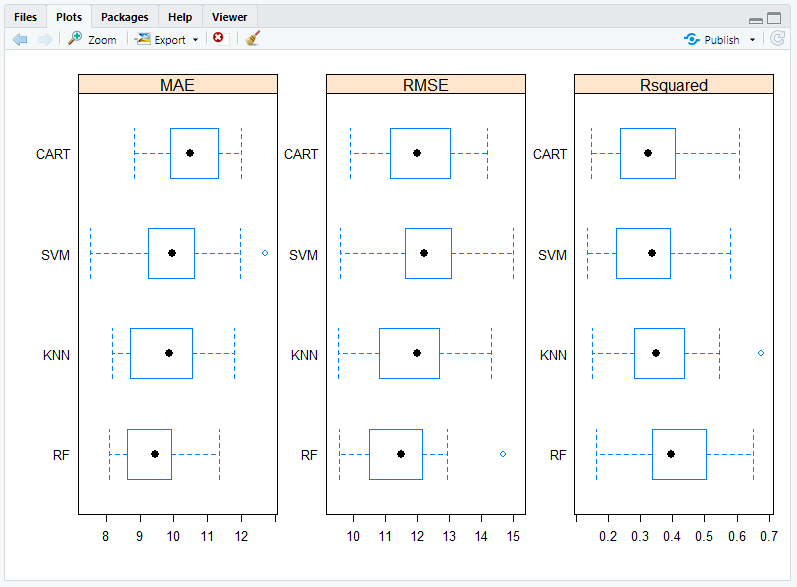
summary(results)



#box wisker

scales <- list(x=list(relation="free"), y=list(relation="free"))

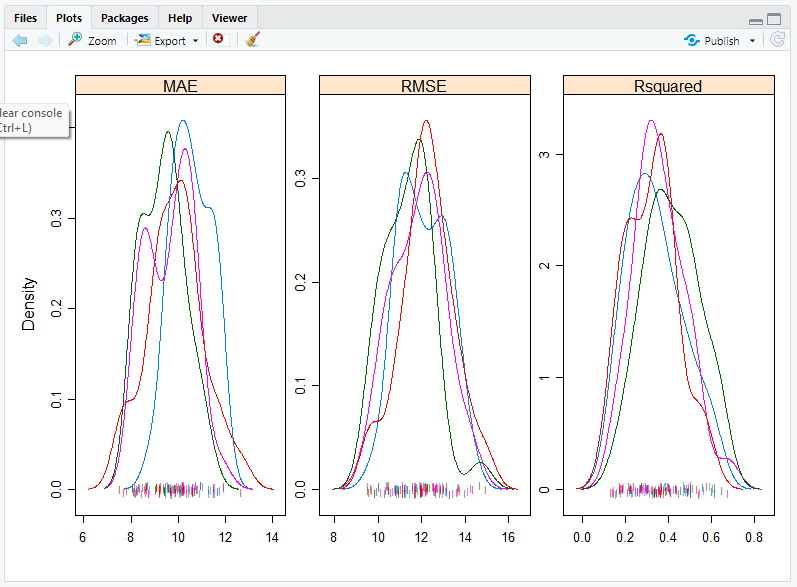
bwplot(results, scales=scales)



#density plots

scales <- list(x=list(relation="free"), y=list(relation="free"))

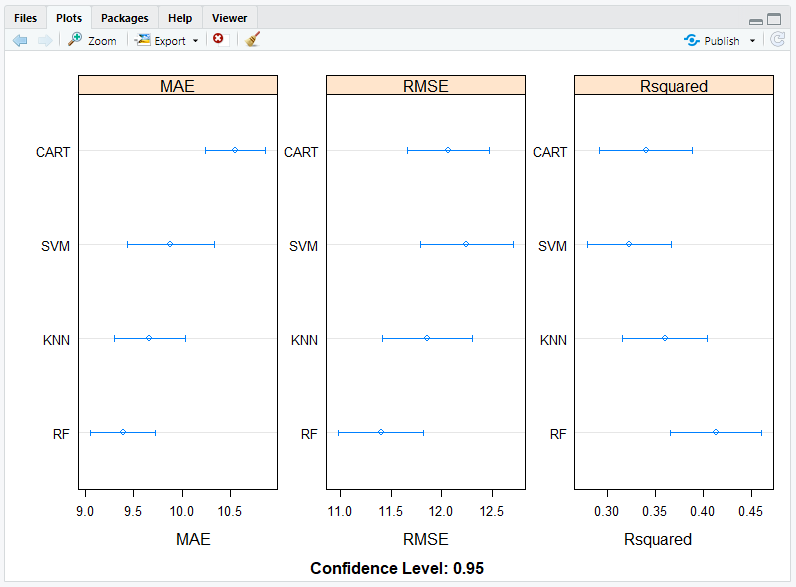
densityplot(results, scales=scales, pch = "|")



#dot plots

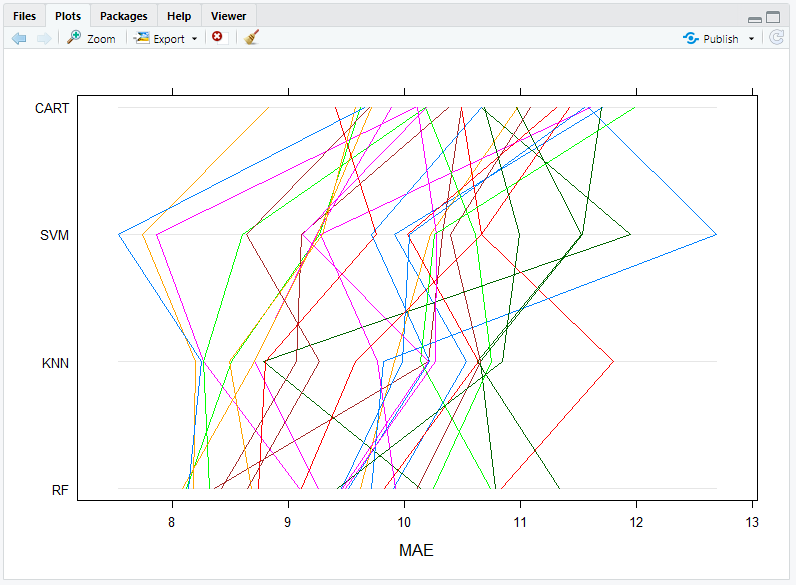
scales <- list(x=list(relation="free"), y=list(relation="free"))

dotplot(results, scales=scales)



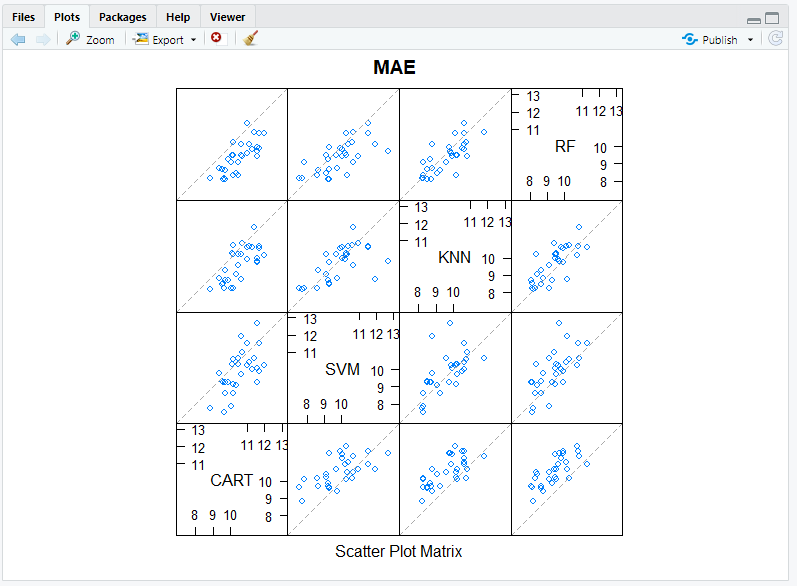
#parallel plots

parallelplot(results)



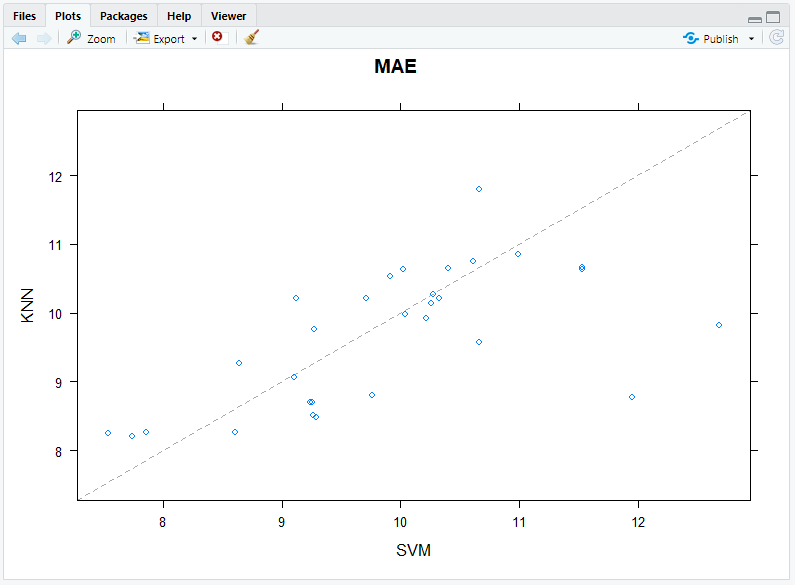
#scatter plot

splom(results)



#pair wise x and y plots

xyplot(results, models=c("KNN", "SVM"))



#statisticall significance test

# difference in model predictions

diffs <- diff(results)

# summarize p-values for pair-wise comparisons

summary(diffs)

