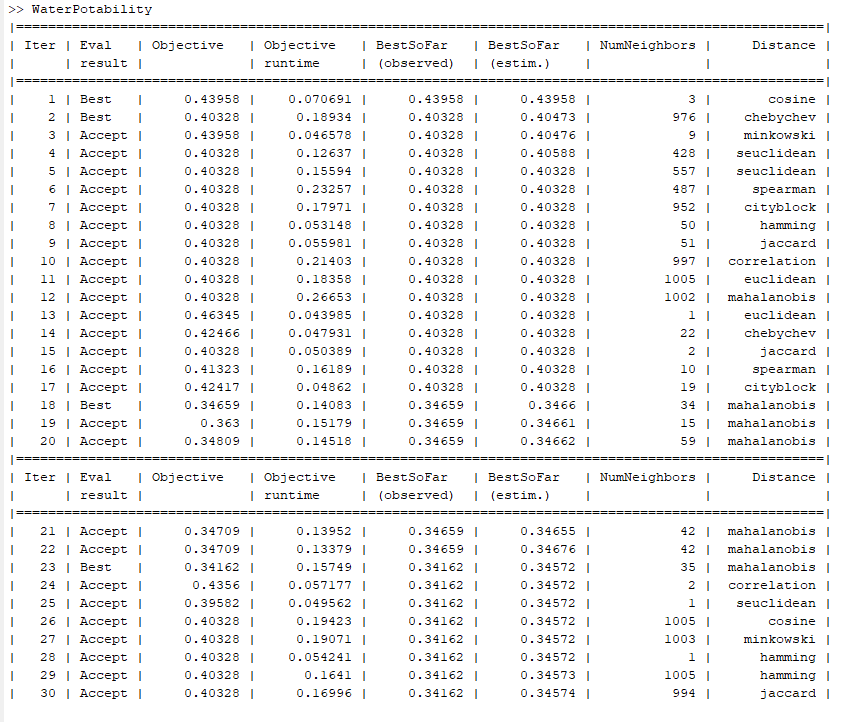
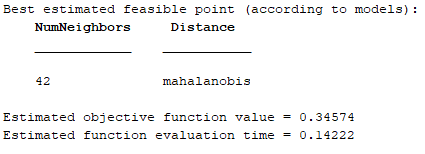
Pengujian akan dilakukan dengan menggunakan K-fold Cross Validation dengan jumlah fold sebanyak 10. Keseluruhan data uji yang digunakan adalah 80 data. keseluruhan data tersebut akan dibagi menjadi 10 partisi. Pengujian dilakukan dengan cara menggunakan salah satu partisi sebagai data uji dan sisanya sebagai data latih. Pengujian akan berulang terus menerus sehingga semua partisi mendapatkan peran sebagai data uji. Tujaun akhir pengujian ini adalah untuk menemukan nilai K terbaik yang dapat digunakan pada model.

Sebelum masuk ke dalam pencarian nilai K, akan dicari terlebih dahulu distance metric dengan nilai terbaik.





Karena distance mahalanobis memiliki rerata nilai yang lebih baik, maka fungsi distance ini yang akan dipilih untuk masuk ke tahap selanjutnya yaitu pencarian jumlah K. Selain mahalonibis, akan dipilih juga distance metric cosin. Dalam pencarian, jumlah K yang dipertimbangkan adalah 5, 15, 35, 41, 43, dan 47.

Berikut adalah hasil pengujiannya:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Distance metric | K | percobaan ke- | | | | | | | | | | K-fold loss | Akurasi |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| cosin | 5 | 0.4125 | 0.3625 | 0.35 | 0.4 | 0.375 | 0.375 | 0.3875 | 0.35 | 0.35 | 0.4 | 0.37625 | 0.62375 |
| cosin | 15 | 0.3375 | 0.3375 | 0.3375 | 0.35 | 0.3375 | 0.375 | 0.4375 | 0.3 | 0.325 | 0.3125 | 0.345 | 0.655 |
| cosin | 35 | 0.4375 | 0.4 | 0.425 | 0.45 | 0.4375 | 0.45 | 0.45 | 0.4625 | 0.425 | 0.4125 | 0.435 | 0.565 |
| cosin | 41 | 0.3875 | 0.4375 | 0.45 | 0.4375 | 0.4 | 0.425 | 0.45 | 0.4625 | 0.475 | 0.4375 | 0.43625 | 0.56375 |
| cosin | 43 | 0.4125 | 0.4875 | 0.4125 | 0.45 | 0.5 | 0.4125 | 0.4375 | 0.425 | 0.425 | 0.5 | 0.44625 | 0.55375 |
| cosin | 47 | 0.425 | 0.4 | 0.525 | 0.4625 | 0.425 | 0.5125 | 0.5125 | 0.4375 | 0.425 | 0.4375 | 0.45625 | 0.54375 |
| mahalanobis | 5 | 0.4125 | 0.35 | 0.4375 | 0.375 | 0.3875 | 0.4125 | 0.35 | 0.35 | 0.35 | 0.3625 | 0.37875 | 0.62125 |
| mahalanobis | 15 | 0.375 | 0.3 | 0.3375 | 0.3875 | 0.325 | 0.325 | 0.425 | 0.3375 | 0.4125 | 0.35 | 0.3575 | 0.6425 |
| mahalanobis | 35 | 0.3875 | 0.45 | 0.4625 | 0.45 | 0.4625 | 0.3875 | 0.5125 | 0.4125 | 0.475 | 0.4625 | 0.44625 | 0.55375 |
| mahalanobis | 41 | 0.55 | 0.5125 | 0.5 | 0.55 | 0.4875 | 0.5 | 0.5 | 0.5 | 0.45 | 0.5125 | 0.50625 | 0.49375 |
| mahalanobis | 43 | 0.475 | 0.525 | 0.4875 | 0.4875 | 0.4625 | 0.5625 | 0.525 | 0.5125 | 0.475 | 0.4875 | 0.5 | 0.5 |
| mahalanobis | 47 | 0.45 | 0.4875 | 0.5 | 0.475 | 0.475 | 0.5125 | 0.5375 | 0.5 | 0.55 | 0.5 | 0.49875 | 0.50125 |

Perbandingan rata-rata nilai akurasi

Dari semua pengujian dengna nilai K dan distance metric berbeda, nilai akurasi tertinggi terdapat pada distance metric cosin dan K=15 dengan tingkat akurasi 65.5% . Nilai akurasi terendah ada pada distance metric mahalanobis dan K=41 dengan tingkat akurasi sebesar 49.375%. Dengan nilai K yang lebih kecil dan memiliki akurasi yang baik, maka dengan demikian nilai K=3 pada distance metric cosin adalah aturan yang paling optimal.

Kesimpulan

Dalam kajian yang dilakukan ini dapat disimpulkan bahwa metode klasifikasi KNN dapat melakukan klasifikasi potensi air untuk dikonsumsi dengan akurasi 65.5% pada nilai K terbaik pada K=15. Fitur yang digunakan diantaranya adalah Ph, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, dan Turbidity.

Pada penelitian ini akurasi sebesar 65.5% menjadi hal yang sangat perlu diperhatikan dan jika memungkinkan dilakukan perbaikan di kemudian hari agar program dapat dengan lebih baik memberikan penilaian apakah air dengan kriteria tertentu dapat dikonsumsi atau tidak.

Gui

