Figure 3.3: Accuracy, Precision and Recall for different distance functions.

Distance Function	Execution Time (K=8)	Execution Time (K=12)	Execution Time (K=15)
Euclidian	0.37400031089782715	0.3698689937591553	0.3779888153076172
Manhattan	0.41501355171203613	0.41440248489379883	0.43491196632385254
Hamming	1.6085307598114014	0.8361709117889404	1.029836654663086

Table 3.2 Execution time for Euclidian, Manhattan and Hamming distance metric

4. Conclusion

Euclidian, Manhattan and Hamming distance functions has been used for the KNN algorithm after implementation it has been observed that KNN algorithm is performing good as comparison Manhattan and Euclidian distance function. However performance all these three function is almost equal as shown in fig 3.3. But in case of execution time KNN algorithm with ecludian distance metric is taking very less time (Table 3.2). So overall, it has been observed that using the Euclidian distance metric is advantageous as comparison to Hamming and Manhattan metric

References

[1] Nakamura, K., Sugawara, M., & Toyama, M. (1991, June). Defect prevention activities and tools. In ICC 91 International Conference on Communications Conference Record (pp. 360-363). IEEE.
[2] Rogerson, S. (2002). The chinook helicopter disaster. IMIS Journal, 12(2).
[3] Murphy, G., & Cubranic, D. (2004, June). Automatic bug triage using text categorization. In Proceedings of the Sixteenth International Conference on Software Engineering & Knowledge Engineering (pp. 1-6). Citeseer..
[4] Anvik, J., Hiew, L., & Murphy, G. C. (2006, May). Who should fix this bug?. In Proceedings of the 28th international conference on Software engineering (pp. 361-370).
[5] Ahsan, S. N., Ferzund, J., & Wotawa, F. (2009, September). Automatic software bug triage system (bts) based on latent semantic indexing and support vector machine. In 2009 Fourth International Conference on Software Engineering Advances (pp. 216-221). IEEE.
[6] Anvik, J., & Murphy, G. C. (2011). Reducing the effort of bug report triage: Recommenders for development-oriented decisions. ACM Transactions on Software Engineering and Methodology (TOSEM), 20(3), 1-35.
[7] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.
[8] Franz, P. (2005). Bayesian Network Classifiers Versus Selective Formula Not Shown-NN Classifier [J]. Pattern Recognition, 38(1), 1-10.
[9] Han, J. (2006). Jiawei Han & Micheline Kamber, Second Edi.
[10] Hand, D. J. (2007). Principles of data mining. Drug safety, 30(7), 621-622.
[11] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 986-996). Springer, Berlin, Heidelberg.
[12] Yong, Z., Youwen, L., & Shixiong, X. (2009). An improved KNN text classification algorithm based on clustering. Journal of computers, 4(3), 230-237.
[13] Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. Journal of machine learning research, 10(2).
[14] Hu, L. Y., Huang, M. W., Ke, S. W., & Tsai, C. F. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus, 5(1), 1-9.
[15] Prasath, V. B., Alfeilat, H. A. A., Hassanat, A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., & Salman, H. S. E. (2017). Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier--A Review. arXiv preprint arXiv:1708.04321.
[16] Kumar, R., Singla, S., Yadav, R. K., & Kumar, D. (2019). An experimental analysis of various data mining techniques for software bug classification. International Journal of Innovative Technology and Exploring Engineering, 8(8 SpecialIssue 3), 108-113.

[17] Kumar, R., & Verma, R. (2012). Classification rule discovery for diabetes patients by using genetic programming. International Journal of Soft Computing and Engineering, 2(4), 183-185.
[18] Kumar, R., Kapil, A. K., & Bhatia, A. (2012). Modified Tree Classification In Data Mining. Global Journal of Computer Science and Technology.

Authors Profile

Raj Kumar is a PhD Research Scholar at Dept. of CSE, I.K. Gujral Punjab Technical University, Jalandhar (Punjab), India. He received his M.Tech degree in Computer Sc. & Engineering from Chaudhary Devi Lal University, Sirsa(Haryana), India in 2007. He obtained his M.Sc. and MCA degree from Kurukshetra University, Kurukshetra, India. He completed his graduation from Kurukshetra University Kurukshetra. His area of interest includes Data Mining and Machine Learning. He has more than 23 referred journal papers in his credit. He has 14 years of teaching experience. Currently he is working as Assistant Professor in the Dept. of CSE at Chandigarh University, Gharuan(Mohali), India

Dr Sanjay Singla completed Ph.D. from Mahrishi Dayanad University, Rohtak and currently working as Professor and Dean at GGS College of Modern Technology, Punjab. His main research work focuses on Vanets, Network Security, Cloud Security and Privacy, Data Mining and IoT. He has more than 20 years of teaching experience.