Separa las anomalías cerca de rama principal del nodo, y esta característica permite construir modelos parciales y efectivos incluso utilizando una poca cantidad de datos para entrenar los datos.

Este método tiene una complejidad lineal por tanto existe poco crecimiento de memoria lo cual es ideal para grandes volúmenes de conjuntos de tasos. Es por esto que, es muy utilizado generalmente en grandes bases de datos.

Las predicciones realizadas por este método pueden converger a un conjunto muy pequeño lo cual permite detectar anomalías de forma eficiente.

Para problemas con muchas variables, e incluso con variables irrelevantes para el modelo; puede crear funciones efectivas para la detección de anomalías incluso con estos atributos adicionales. Incluso trabaja bien si no existen anomalías presentes en el conjunto de datos.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining* (pp. 413-422). IEEE.

Desventajas K-means Detección de anomalías

En los métodos de aprendizaje supervisado es necesario crear clasificadores utilizando datos de prueba y test. Para detección de anomalías es mejor considerar que la información siempre esa segada. Por tanto, debe ser mejor una mayor cantidad de datos con éxito y unas pocas anomalías; esto generalmente es un problema para algoritmo de clasificación como clúster jerárquico.

K-means generalmente puede manjar mejor datos numéricos y por tanto los datos pueden estar segados en caso de no estar normalizados.

A pesar de definir un buen valor de k puede existe otro problema. Si en el futuro existen nuevos datos, algo muy común en escenarios de detección de anomalías entonces resulta necesario cambiar el valor de K. AL igual que los datos, este valor puede cambiar y periódicamente será necesario buscar un valor de K óptimo.

<https://www.inovex.de/blog/disadvantages-of-k-means-clustering/>

Nota: Para el presente trabajo no se realizará la separación de datos en conjunto de prueba y test, considerando que en la detección de anomalías generalmente se buscan los valores atípicos. Debido a que las técnicas de agrupamiento son no supervisadas, se va a formar y revisar si los clústeres resultantes coinciden con estas etiquetas Esto en función de las consultas realizadas en clase y en el foro.