**Icono

Descripción generada automáticamenteImagen que contiene Círculo

Descripción generada automáticamente**

**Universidad de Montemorelos**

**Catedrático: Harvey Alférez, PhD.**

**Alumnos: Javier Ramón González, Geovanni Dzul Escobar**

**Asignatura: Ciencia de Datos**

**Actividad: Documentación.**

**Séptimo semestre**

**Resultados de los clasificadores**

Para la creación de los modelos clasificatorios utilizamos la librería de scikit-learn de Python, la cual nos facilita la creación e implementación de estos modelos para el análisis y predicción de valores con nuestros datos. Para este modelo se utilizaron datos de automóviles, el dataset contiene 6 variables independientes y 1 dependiente, las variables son: Precio, Costo de mantenimiento, Numero de puertas, Numero de personas, Tamaño de cajuela, Seguridad; La variable de salida nos indicaría dependiendo del valor de cada una de las variables anteriores, si el carro es poco aceptable, aceptable, bueno o muy bueno para su compra.

Tabla

Descripción generada automáticamentePara el entrenamiento y predicción utilizamos 4 clasificadores y 1 red neuronal, los clasificadores fueron: Knn, Decision Tree, Support Vector Machine y Logistic Regression y para la red neuronal utilizamos Feedforward Backpropagation. Los datos fueron entrenados con cada uno de estos elementos y medimos su funcionalidad y su exactitud con base a 4 variables: Precision, Recall, F1-Score y Accurancy.

Como se puede observar en la imagen anterior se realizó una comparación entre cada uno de los modelos predictivos, para poder escoger el modelo que tenga valores mas exactos a la hora de predecir en ese caso si el auto es poco aceptable, aceptable, bueno o muy bueno, se decidió mantener las 4 clases ya que esto ayuda a dar un resultado más exacto y también da posibilidad a puntos intermedios. Como se puede observar cada modelo fue sustentado con respecto a 4 clasificaciones ya antes mencionadas y se realizo una comparativa entre ellos para escoger cual es el mejor modelo. Al verificar y comparara Tabla

Descripción generada automáticamenteestos modelos pudimos encontrar que el mejor modelo fue el de Decision Tree.

Como se observa este clasificador nos da un Accurancy de 0.97 lo que seria equivalente a 97% de exactitud, la exactitud mide el porcentaje de casos que el modelo ha acertado. Esta métrica nos ayuda a darnos una idea de la eficacia de nuestro modelo, mas no debería ser concluyente a la hora de la comparación, ya que puede llegar a predecir de una manera equivocada el modelo, es por eso por lo que no solo se tomo en cuenta el accurancy para escoger este modelo sino también las otras 3 variables ya mencionadas, a continuación, se muestra la comparación del modelo escogido con respecto a los otros, además de explicar también cada una de las variables.

Tabla

Descripción generada automáticamente**Decision Tree VS KNN:**

Tabla

Descripción generada automáticamente

En las 2 imágenes anteriores podemos observar la comparación del modelo escogido contra Knn el cual es otro modelo de clasificación, como podemos observar, el modelo escogido tiene valores lo suficiente altos y no tan variables en cada una de las clases con respecto a la precisión, pero que es la precisión: Es una métrica para medir la calidad del modelo, eso quiere decir que mide cada clase y da la precisión de acierto que tendrá para cada una, como observar en KNN hay algunas clases las cuales tienen una precisión de 0 mientras que en el modelo escogido tenemos la mayoría con porcentajes similares y no tan variables, también podemos compara la variable de Recall, esta variables nos muestra la capacidad de identificar la clase, en el caso de la primera clase para KNN podemos observar que la capacidad de que tiene para identificar esa clase es muy baja, al igual que para la segunda y tercera; pero si observamos al modelos escogido, podemos observar que la capacidad que tiene para identificar cada clase es suficiente buena para poder escoger la mejor. Por último, comparamos el valor F1-score, este valor combina las medidas de precisión y recall en un solo valor. Lo cual nos da una idea de como es el rendimiento tanto de precisión como de recall. Este valor concluye que tanto precisión como recall son importantes. En este caso en KNN tenemos valores muy bajos para las primeras tres clases, mientras que para la ultima el valor es mayor, pero en el clasificador escogido podemos ver como los porcentajes no están tan separados, lo cual nos indica que el modelo tiene tanto un valor de precisión bueno como de Recall.

Ya explicadas las variables que comparamos, seguiremos mostrando las comparaciones con los otros 3 modelos que faltan.

Tabla

Descripción generada automáticamente**Decision Tree VS Logistic Regresion:**

Tabla

Descripción generada automáticamente

Como ya vimos anteriormente el accurancy no debería ser el único valor que comparáramos, pero aquí podemos observar un gran cambio, ya que el accurancy para Logistic regresión es del 77% mientras que para Decision Tree es de 97%. Además, podemos ver que para algunas variables y clases los porcentajes tienen una gran variabilidad, en el apartado de recall para la clase Bueno el porcentaje es de 0.53, mientras que para las otras es mayor, por lo cual esta muy alejado de los porcentajes de las otras clases, lo cual podría darnos falsos positivos o falsos negativos a la hora de querer predecir esa clase, para este tipo de casos podríamos decidir quitar dicha clase para poder trabajar solo con las que su puntuación no varia tanto, pero si observamos al clasificador escogido podemos ver que sus porcentajes no son tan variables por lo que utilizar este clasificador no es adecuado.

**Decision Tree VS Support Vector Machine:**

Tabla

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

En esta comparación lo que podemos observar es que para las clases la mayoría tienen valores muy bajos para cada una de las variables, lo cual nos indican el mal funcionamiento de este clasificador, aunque su accurancy no es tan bajo como el anterior, las otras métricas si lo son y esto podría afectar a la hora de la predicción de datos.

**Decision Tree VS Feedforward Backpropagation:**

Tabla

Descripción generada automáticamente

Tabla

Descripción generada automáticamente

Aunque feedforward Backpropagation no sea en si un clasificador, también lo comparamos con respecto a nuestro clasificador seleccionado. Feedforward Backpropagation es una red neuronal, la cual también entrenamos para la predicción de datos, como podemos observar en la comparación esta red neuronal es la que más competencia le hace a el clasificador que nosotros seleccionamos ya que podemos observar que su accurancy es del 91%, lo cual no esta tan alejado del 97%, pero al observar la diferencia que hay en las otras variables, podemos concluir que tampoco utilizar esta red neuronal seria lo mejor, ya que sus porcentajes para cada una de las clases con respecto a las variables, en algunos casos, son o muy bajos o tienen una separación considerable.

**Conclusión:**

Después de comparar todos los clasificadores entre ellos concluimos que la mejor decisión seria utilizar Decision Tree para la clasificación y predicción de nuestros modelos no solo por el nivel de Exactitud, sino por el buen desempeño en las demás variables, por lo que Decision Tree es el clasificador que se utilizara para la predicción de valores