



Tecnológico de Monterrey

Campus Estado de México

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Actividad 5. Regresión logística

Juan Pablo Castañeda Serrano

A01752030

Estadística descriptiva y coeficiente de correlación:

Las variables Lag1 a Lag5 parecen seguir una distribución cercana a 0. El volumen ha aumentado a lo largo de los años, lo que indica una correlación positiva entre el año y el volumen. Las correlaciones entre las variables Lag y Today son bajas, lo que indica que los rendimientos de las semanas anteriores no están fuertemente correlacionados con los rendimientos actuales.

Modelo logístico completo:

El único predictor significativo al nivel del 5% es Lag2 con un valor p de 0,0296. Esto sugiere que el desempeño de dos semanas tiene algo que ver con la dirección actual. Un coeficiente positivo en Lag2 sugiere que un mayor rendimiento hace dos semanas está asociado con una mayor probabilidad de una tendencia "alcista" en la semana actual.

Modelo logístico simplificado:

Aunque Lag2 es el único predictor significativo en el modelo completo, su importancia en el conjunto de entrenamiento es marginal (valor p = 0,06554). El volumen no es significativo en el modelo.

Visualización:

El gráfico debe mostrar la relación general entre Lag2 y la probabilidad prevista, coloreada según la dirección verdadera. Esto puede proporcionar una indicación visual de cómo se ajusta el modelo a los datos y si existen patrones claros. Evaluación del modelo:

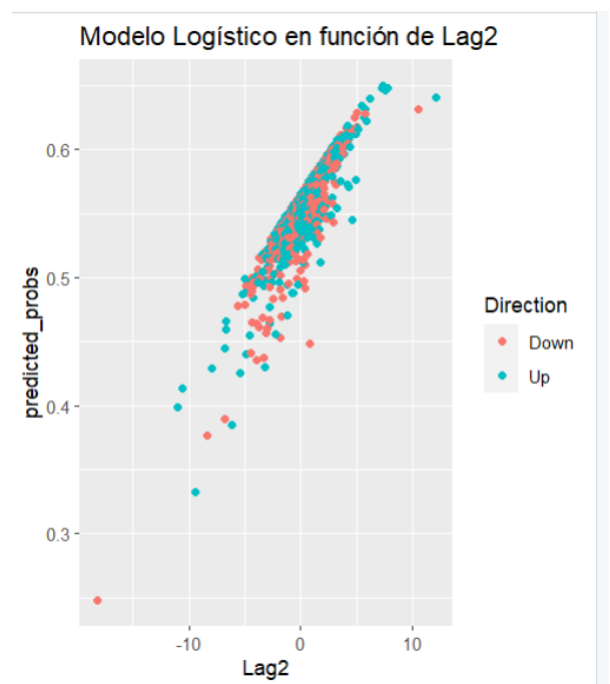
La matriz de confusión muestra que el modelo tiene dificultades para predecir correctamente la dirección. De 43 semanas, el modelo "Abajo" predice correctamente 20. De 61 semanas, el modelo "Arriba" predice correctamente 36. La prueba de chi-cuadrado no muestra diferencias significativas entre las predicciones y las observaciones reales, lo que indica que el modelo no es mejor que la selección aleatoria.

Interpretación en el contexto del problema:

El modelo logístico basado en Lag2 y Volumen no parece especialmente eficaz para predecir la dirección del mercado de valores. Un coeficiente positivo en Lag2 sugiere que un mayor rendimiento hace dos semanas puede estar asociado con una mayor probabilidad en la semana actual, aunque esta asociación es marginalmente significativa.

El hecho de que el modelo no sea muy eficiente puede deberse a la naturaleza inherentemente volátil y compleja del mercado de valores, donde múltiples factores externos pueden afectar los movimientos semanales. Además, utilizar información pasada para predecir movimientos futuros es una estrategia común, pero no siempre efectiva, como muestra este modelo.

Aunque Lag2 parece tener un efecto, es importante recordar que la correlación no implica causalidad. Es posible que otros factores que no están presentes en la serie de datos estén influyendo en la dirección del mercado.



Conclusión

En conclusión, aunque el modelo puede proporcionar alguna idea sobre la relación entre las ganancias quincenales y la dirección actual del mercado, no se debe confiar en él para hacer predicciones precisas sobre el mercado. Para mejorar la precisión del modelo, sería recomendable explorar otros métodos o agregar más variables que puedan afectar la dirección del mercado.