15-06-2021

Ensemble, tipos:

Secuencial: Votación.- utiliza la moda el hard voting. El soft voting trabaja con la media. Si es regresión nos quedamos con la media de los diferentes modelos de regresión

Distintos modelo,

Modo paralelo: Mismo modelo: bagging (distintas muestras con reemplazamiento) y pasting (distintas muestras sin reemplazamiento)

Bootstrap.- se busca que los estimadores sean lo más independientes posible

Bagging, cómo funciona

Boosting.- se ejecuta un algoritmo y con este entrenado vamos a entrenar el siguiente modelo y así sucesivamente

Adabost.- recoge el entrenamiento del anterior modelo

Grid search.-

Pipeline:

le damos los hiperparámetros que queremos para hacer el entrenamiento (quitar columnas, estandarizar, seleccionar las 20 columnas que mejor predicen, …)

Permite probar diferentes modelos con diferentes hiperparámetros, todos a la vez

cv.- cross validation

Simpleimputer.- asigna a los valores missing lo que le diga (una media, etc)

chi2 se le dice que se quede con las x columnas que mejor correlacionen. Nosotros le damos el x

njob=-1; que pruebe todos los núcleos del procesador

simpleImputer por defecto es la media

standarized

16-06-2021

Taller Pensamiento analítico

Tipos de sesgos colectivos:

* Selección.- Sistemas de recolección de datos .
  + De confirmación.- Ignorar deliberadamente evidencia que no se ajusta a una conclusión deseada o preconcebida
  + De supervivencia.- Concentrarse en las personas o cosas que superaron un proceso de selección pasando por alto a aquellas que no lo hicieron. Ignora los fracasos
  + De anclaje.- Confiar demasiado en una parte de la información e interpretar las conclusiones del análisis en torno a esos hallazgos
  + Ausencia de evidencia.- No abordar el impacto de la ausencia de información en las conclusiones analíticas
  + Sesgo de novedad.- darle más peso a la información más reciente que la que tiene más tiempo
  + Sesgo de volumen de información.- Más información no siempre significa una mejor decisión. Sumamos datos que no influyen en el objeto de estudio añadiendo ruido al análisis
  + Sesgo de observación.- Tendencia a buscar en lugares donde se espera obtener buenos resultados o donde es muy conveniente observar
  + Sesgo de financiación.- Tendencia inconsciente a sesgar los modelos, los datos o la interpretación de los datos de forma que favorezcan los objetivos de un patrocinador financiero o empleador
  + Sesgo de muestreo.- La selección de los grupos o de los datos no es representativa(no intencionadamente) del objeto de análisis. Ley de los números pequeños
* Proceso.- Análisis y manejo de datos
  + Sesgo de inducción.- Estimular, indirecta o deliberadamente la respuesta ya en la formulación de la pregunta
  + Sesgo de confianza.- Tener demasiada confianza en nuestras capacidades nos lleva a asumir mayores riesgos y cometer más errores
  + Apofenia.- Sobreestimación del valor de los patrones percibidos en los datos aleatorios. Correlaciones irrelevantes y Falsa Causalidad
  + Sesgo de encuadre.- La forma en que se presentan los datos puede influir en su percepción
  + Sesgo de disponibilidad.- Si algo se puede recordar, debe ser importante, o al menos más importante que los sucesos alternativos que no se recuerdan tan fácilmente
  + Sesgo de escalada irracional.- Justificar el aumento de la inversión de tiempo basándose en la inversión de tiempo anterior y existente
* Sociales.- Toma de decisión, storytelling
  + Pensamiento grupal. Elegir la opción con la que está de acuerdo la mayoría del grupo o ignorar los conflictos dentro del grupo debido al deseo de lograr un consenso
  + Efecto halo.- Aceptar decisiones de personas a las que admiras sin cuestionarlas. O el efecto halo inverso, alguien no te agrada, por lo tanto, no puede tener razón
  + Sesgo de espejo.- Suponer que otros actuarán igual que nosotros al encontrarse en circunstancias similares
  + Estereotipo.- Esperar que un grupo o persona tenga ciertas características sin tener información real sobre la persona o grupo
  + Maldición del conocimiento.- Al comunicarnos con otros individuos, suponer, sin saberlo, que los demás tienen los conocimientos necesarios para entender lo que se dice
  + Punto ciego.- Notar los sesgos cognitivos y motivacionales mucho más en otros que en uno/a mismo/a

Soluciones:

* Habilidades
* Tener tiempo
* Enfoque
* Técnicas
  + Análisis DAFO

Estrategias de mitigación

* Sesgos de selección
  + Utilidad de los datos
  + Utilidad de las fuentes
  + Credibilidad de los datos
  + Lagunas de información
* Sesgos de proceso
  + Técnica de los seis sombreros
* Sesgos sociales
  + Hipótesis alternativa
  + Hipótesis contrapuestas
  + Abogado del diablo
  + Mapeo lógico

Hábitos de los analistas expertos:

1. Cuestionar los supuestos clave
2. Considerar hipótesis alternativas
3. Buscar datos inconsistentes que proporcionen justificación
4. Concentrarse en los impulsores clave que mejor explique lo que ha ocurrido o lo que está por suceder