

Data Analytics: práctica final

Descripción de la práctica

El objetivo de esta práctica es la creación de un modelo completo de análisis de datos, que incluya todos los pasos de la metodología CRISP-DM, y basado en el dataset proporcionado a cada grupo. La entrega consistirá en:

- Un documento, Word o PDF, de descripción del modelo, y que contenga:
 - Business Plan;
 - Development Plan; el análisis a entregar tiene que contener los pasos habituales de una tarea de análisis de datos de este tipo, entre otros:
 - Análisis inicial y pre-procesado de los datos;
 - Selección de las features significativas;
 - Entrenamiento de distintos modelos;
 - Comparación de modelos por medio de validación cruzada;
 - Conclusiones acerca de la mejor solución encontrada.Los modelos tendrán que incluir toda la información necesaria para su reproducción, como por ejemplo parámetros, algoritmos utilizados, etc.
 - Instrucción para el despliegue del modelo, o manual de usuario del sistema desplegado. *Este tercer punto es opcional.*
- Cualquier fichero relacionado con el desarrollo del modelo, y que sea necesario para evaluar el trabajo realizado. Por ejemplo, esto incluye:
 - Ficheros STR de SPSS, KNWF/KNAR de KNIME, etc.
 - Cualquier código (en Python, C, Java, etc.) que se haya desarrollado para pre-procesar los datos, enriquecerlos, etc.
 - En el caso de añadir dichos ficheros, es también recomendable añadir un fichero adicional con la explicación del contenido de cada uno de ellos.

Nota importante: cualquier resultado presentado tendrá que ser justificado, ya sea por medio de un modelo de SPSS o KNIME, bien por medio de un código específico. Cualquier resultado no soportado no será tenido en cuenta de cara a la evaluación.

La práctica completa será entregada subiendo un fichero ZIP al sistema Moodle. La fecha final para entregar el material es el **Jueves 17 de Mayo a las 22:00**. Adicionalmente, cada grupo tendrá que presentar los resultados más importantes que se han obtenido en las dos clases de la semana del 21/24 de Mayo (máximo de 10 minutos para cada grupo).

Descripción del problema

Un centro de investigación clínica está estudiando la posibilidad de desplegar un modelo de análisis de datos que ayude a gestionar pacientes que padecen problemas derivados de la diabetes. En concreto, se quiere desarrollar un sistema capaz de predecir si un paciente, que ha sido ingresado y dado de alta, tiene alta probabilidad de sufrir recaídas.

Para ello, se ha recolectado información de 10 años (1999-2008) de cuidados clínicos en 130 hospitales en Estados Unidos, incluidos datos acerca de los pacientes y de las pruebas realizadas. A continuación, se han retenido aquellos pacientes que han satisfecho los siguientes criterios:

- El paciente ha sido admitido en el hospital, y ha sido diagnosticado de alguna forma de diabetes.
- El paciente ha sido ingresado por un mínimo de un día, y por un máximo de 14.
- Algún test de laboratorio ha sido ejecutado durante la estancia.
- Alguna medicación ha sido prescrita.

Entre las features disponibles en el dataset, se incluyen:

- Datos del paciente, cuales edad, sexo o peso.
- Tipo de ingreso – ver tabla adjunta.
- Destino del alta – ver tabla adjunta.
- Procedencia – ver tabla adjunta.
- Valores de diagnósticos – no se especifica a que pruebas corresponden.
- Presencia de alteraciones en distintos marcadores.
- “readmitted”: “YES” si el paciente ha vuelto a visitar el hospital después de este día, “NO” en caso contrario. Esta es la variable a predecir por el modelo.

Nota: si una feature no está descrita en el listado anterior, su significado no ha podido esclarecerse con el centro de investigación. Sin embargo, esto no impide su uso.

Nota: no todas las features están disponibles para todos los grupos.

admission_type_id:

- 1, Emergency
- 2, Urgent
- 3, Elective
- 4, Newborn
- 5, Not Available
- 6, NULL
- 7, Trauma Center
- 8, Not Mapped

discharge_disposition_id:

- 1, Discharged to home
- 2, Discharged/transferred to another short term hospital
- 3, Discharged/transferred to SNF
- 4, Discharged/transferred to ICF
- 5, Discharged/transferred to another type of inpatient care institution
- 6, Discharged/transferred to home with home health service
- 7, Left AMA
- 8, Discharged/transferred to home under care of Home IV provider
- 9, Admitted as an inpatient to this hospital
- 10, Neonate discharged to another hospital for neonatal aftercare
- 11, Expired
- 12, Still patient or expected to return for outpatient services
- 13, Hospice / home
- 14, Hospice / medical facility
- 15, Discharged/transferred within this institution to Medicare approved swing bed
- 16, Discharged/transferred/referred another institution for outpatient services
- 17, Discharged/transferred/referred to this institution for outpatient services
- 18, NULL
- 19, "Expired at home. Medicaid only, hospice."
- 20, "Expired in a medical facility. Medicaid only, hospice."
- 21, "Expired, place unknown. Medicaid only, hospice."
- 22, Discharged/transferred to another rehab fac including rehab units of a hospital .
- 23, Discharged/transferred to a long term care hospital.
- 24, Discharged/transferred to a nursing facility certified under Medicaid but not certified under Medicare.
- 25, Not Mapped
- 26, Unknown/Invalid
- 30, Discharged/transferred to another Type of Health Care Institution not Defined Elsewhere
- 27, Discharged/transferred to a federal health care facility.

- 28, Discharged/transferred/referred to a psychiatric hospital of psychiatric distinct part unit of a hospital
- 29, Discharged/transferred to a Critical Access Hospital (CAH).

admission_source_id:

- 1, Physician Referral
- 2, Clinic Referral
- 3, HMO Referral
- 4, Transfer from a hospital
- 5, Transfer from a Skilled Nursing Facility (SNF)
- 6, Transfer from another health care facility
- 7, Emergency Room
- 8, Court/Law Enforcement
- 9, Not Available
- 10, Transfer from critical access hospital
- 11, Normal Delivery
- 12, Premature Delivery
- 13, Sick Baby
- 14, Extramural Birth
- 15, Not Available
- 17, NULL
- 18, Transfer From Another Home Health Agency
- 19, Readmission to Same Home Health Agency
- 20, Not Mapped
- 21, Unknown/Invalid
- 22, Transfer from hospital inpt/same fac reslt in a sep claim
- 23, Born inside this hospital
- 24, Born outside this hospital
- 25, Transfer from Ambulatory Surgery Center
- 26, Transfer from Hospice