

Curso de Aprendizaje Máquina, Otoño 2024

Examen II

1. El papel del aprendizaje automático en la supervivencia de las fintech

El ecosistema fintech incluye varios tipos de empresas. Si bien las empresas emergentes que desarrollan nuevas tecnologías o servicios financieros suelen asociarse con el sector fintech, también abarca a empresas tecnológicas establecidas que se expanden hacia los servicios financieros y a instituciones financieras tradicionales que adoptan nuevas tecnologías. Este panorama diverso ha dado lugar a innovaciones en múltiples sectores financieros, incluidos la banca, los seguros, la inversión y los sistemas de pago. Las aplicaciones fintech abarcan una amplia gama de servicios financieros. Entre ellos se incluyen la banca digital, los pagos móviles y las billeteras digitales, las plataformas de préstamos entre pares, los asesores robotizados y el comercio autónomo, la tecnología de seguros, la cadena de bloques y las criptomonedas, la tecnología regulatoria y las plataformas de financiación colectiva.

Las empresas fintech, a pesar de revolucionar la industria financiera, también enfrentan desafíos significativos. Dos de sus problemas más apremiantes son: el riesgo de otorgar crédito al consumidor y la pérdida de clientes.

1.1. Riesgo de otorgar crédito al consumidor

A medida que las fintech extienden crédito a un rango más amplio de clientes, a menudo en base a historiales crediticios limitados, evaluar con precisión la solvencia crediticia de los consumidores se vuelve crucial. Las evaluaciones incorrectas del riesgo existente al otorgar créditos pueden llevar a un aumento de los incumplimientos de préstamos, dañando la salud financiera de la empresa.

1.2. Pérdida de clientes

En un mercado altamente competitivo como el Mexicano, la retención de clientes es esencial. Anticipar la pérdida de clientes permite a las fintech abordar proactivamente los problemas e implementar estrategias de retención. Una alta pérdida de clientes puede erosionar los ingresos y dañar la reputación de la empresa. La pérdida de clientes puede deberse a diversas causas, por ejemplo: el mal servicio brindado por la institución, la calidad de los productos ofertados por la competencia o la difusión de información falsa. Un caso reciente del sector bancario es la pérdida de clientes que sufrió Banco Azteca debido a la propagación de noticias falsas durante el primer semestre de 2024.

Para abordar estos desafíos, las fintech están recurriendo cada vez más a la IA, particularmente al desarrollo de soluciones basadas en técnicas de aprendizaje automático. Al aprovechar eficazmente las nuevas herramientas y servicios basados en aprendizaje automático, las fintech pueden mitigar riesgos, mejorar las experiencias de los clientes e impulsar su crecimiento sostenible.

2. Objetivo de la actividad

WelfareBnk es una fintech que incursionó hace cinco años en el mercado Mexicano (adquirió una institución financiera que tenía cincuenta años en el país). Actualmente, la institución tiene diversos problemas relacionados con la cartera vencida y la pérdida de clientes, por lo que ha solicitado tu empresa de consultoría el desarrollo de dos servicios/aplicaciones que le ayuden a:

- Identificar clientes que potencialmente abandonarán el banco. Esto les permitirá anticiparse y ofrecer a los clientes incentivos para no abandonar la institución.

- Reducir la cartera vencida, mediante una mejor evaluación de riesgos sobre la solicitudes de créditos realizadas por los clientes.

Como ingeniero con experiencia en aprendizaje automático se te ha asignado la construcción, entrenamiento y validación de modelos que puedan ser utilizados como parte de los servicios que ayudarán a la empresa a evaluar mejor los riesgos de crédito al consumidor y la identificación de clientes que potencialmente abandonarán la institución.

3. Disponibilidad de datos

Debido a temas de privacidad y confidencialidad de la información, WelfareBnk ha compartido con tu empresa consultora únicamente una muestra representativa de datos. La muestra contiene la siguiente información sobre clientes que decidieron abandonar la institución y clientes que han permanecido en la misma:

- Identificador del cliente
- La puntuación crediticia del cliente
- La localidad de residencia del cliente
- Género
- Edad
- Permanencia del cliente en la institución
- Saldo promedio mensual
- Número de productos que tiene el cliente
- Si tiene tarjeta de crédito o no
- Si el estatus de cliente es activo o no
- El salario estimado
- Si el cliente abandonó o no la institución

En la figura 1 se muestran ejemplos de los registros que se tienen con información relacionada a la perdida y retención de clientes en WelfareBnk.

	id_cliente	puntuacion_crediticia	Localidad	Genero	Edad	permanencia	saldo	num_de_productos	tiene_tarjeta_credito	es_cliente_activo	salario_estimado	abandono
0	15634602	619	Guadalajara	Femenino	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	CDMX	Femenino	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	Guadalajara	Femenino	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	Guadalajara	Femenino	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	CDMX	Femenino	43	2	125510.82	1	1	1	79084.10	0

Figura 1: Registros con información asociada a la perdida y retención de clientes en WelfareBnk

Adicionalmente, la muestra de datos incluye información relacionada a valoraciones de riesgo basadas en el perfil de los clientes (realizadas por expertos) sobre el otorgamiento de créditos:

- Edad del cliente
- Ingresos anuales
- Propiedad de la vivienda (alquiler, propia o hipotecada)
- Antigüedad en el empleo actual (en años)

- Intención de préstamo (personal, educación, médico, emprendimiento, consolidar deudas, mejora del hogar)
- Grado del préstamo (A, B,..., F), en donde A es la mejor calificación e indica la capacidad para cumplir con sus obligaciones financieras a tiempo.
- Monto del préstamo
- Tasa de interés
- Estado crediticio (0 no es moroso, 1 es moroso)
- Porcentaje de ingresos
- Incumplimiento histórico (Y, N)
- Duración del historial crediticio (en años)

En la figura 2 se muestran algunos ejemplos de registros con la información que se tiene disponible sobre la evaluación de riesgos crediticios de los clientes en WelfareBnk.

	edad	ingreso_anual	propiedad_vivienda	antiguedad_empleo	intencion_credito	grado_credito	monto_credito	tasa_interes	estado_credito	porcentaje_de_ingreso	incumplimiento_historico	historial_credicio
0	22	59000	alquiler	123.0	personal	D	35000	16.02	1	0.59	Y	3
1	21	9600	propia	5.0	educacion	B	1000	11.14	0	0.10	N	2
2	25	9600	hipoteca	1.0	medico	C	5500	12.87	1	0.57	N	3
3	23	65500	alquiler	4.0	medico	C	35000	15.23	1	0.53	N	2
4	24	54400	alquiler	8.0	medico	C	35000	14.27	1	0.55	Y	4

Figura 2: Registros con información asociada a riesgos crediticios de los clientes en WelfareBnk

Para descargar las muestras de datos, acceda al curso de Aprendizaje Máquina (O2024_ES/3178G) que se encuentra en CANVAS. El enlace de descarga se encuentra habilitado en la sección correspondiente a la semana 15 del curso.

4. Métricas de desempeño

La evaluación final de sus modelos se realizará tomando en consideración la exactitud y precisión con respecto a las variables objetivo sobre los conjuntos de datos de prueba. Es importante considerar que estos datos de prueba no se incluyen en la muestra de datos que tiene a su disposición para el desarrollo de la actividad. Por lo anterior, es importante cuidar que mientras se logra un buen desempeño en términos de exactitud y precisión, se mantenga una buena capacidad de generalización en sus modelos.

5. Evaluación

La puntuación total obtenida por el desarrollo de la actividad se calculará mediante,

$$puntos_modelo_riesgo = \frac{exactitud_puntos + precision_puntos}{2}$$

$$puntos_modelo_abandono = \frac{exactitud_puntos + precision_puntos}{2}$$

$$Total = (0.5 * puntos_modelo_riesgo) + (0.5 * puntos_modelo_abandono)$$

En las tablas 2 y 1 se indica la cantidad de puntos que obtendrá por su modelo de estimación de riesgo crediticio acorde al desempeño logrado en términos de la exactitud y precisión sobre los datos de prueba. Mientras que las tablas 4 y 3 indica la cantidad de puntos que obtendrá por su modelo de predicción del abandono de clientes acorde al desempeño logrado en términos de la exactitud y precisión sobre los datos de prueba.

Precisión del modelo	Puntos
mayor o igual a 80 %	100
mayor o igual a 70 %	90
mayor o igual a 60 %	80
mayor o igual a 55 %	70
mayor a 50 %	60

Tabla 1: Puntos obtenidos para la evaluación acorde a la precisión lograda por el modelo que predice el riesgo crediticio de clientes sobre datos de prueba

Exactitud del modelo	Puntos
mayor o igual a 90 %	100
mayor o igual a 80 %	90
mayor o igual a 70 %	80
mayor o igual a 60 %	70
mayor a 50 %	60

Tabla 2: Puntos que se obtienen para la evaluación acorde a la exactitud lograda por el modelo que predice el riesgo crediticio de clientes sobre datos de prueba

Precisión del modelo	Puntos
mayor o igual a 80 %	100
mayor o igual a 70 %	90
mayor o igual a 60 %	80
mayor o igual a 55 %	70
mayor a 50 %	60

Tabla 3: Puntos obtenidos para la evaluación acorde a la exactitud lograda por el modelo que predice el riesgo crediticio de clientes sobre datos de prueba

Exactitud del modelo	Puntos
mayor o igual a 80 %	100
mayor o igual a 70 %	90
mayor o igual a 65 %	80
mayor o igual a 60 %	70
mayor a 55 %	60

Tabla 4: Puntos que se obtienen para la evaluación acorde a la exactitud lograda por el modelo que predice el riesgo crediticio de clientes sobre datos de prueba

- Si construyes tus modelos utilizando **TensorFlow** y **Keras**, se sugiere almacenar sus modelos en formato: **keras**. Para este formato, consulte la documentación en el siguiente enlace:
(https://keras.io/guides/serialization_and_saving/)
- Si construyes tus modelos utilizando **sklearn**, se sugiere almacenar sus modelos en formato: **skops**. Para este formato, consulte la documentación de su API en el siguiente enlace:
https://scikit-learn.org/stable/model_persistence.html#using-skops

Adicionalmente, puede optar por utilizar otros formatos para almacenar sus modelos entrenados. Sin embargo, independientemente del formato que decida utilizar para almacenar los modelos entrenados, verifique que se almacenen de forma correcta y que pueden ser cargados sin problema para realizar predicciones con ellos.

Productos entregables

Con el fin de realizar la evaluación de su actividad es indispensable que entregue mediante CANVAS lo siguiente:

- Código fuente de la implementación de sus soluciones en archivos .py o .ipynb.
- Su mejor modelo entrenado para estimación del riesgo crediticio y un archivo .py que contenga una función que permita preparar datos de entrada para su modelo entrenado y una función que ayude a interpretar la salida de su modelo.
- Su mejor modelo entrenado para la predicción de abandono de clientes y un archivo .py que contenga una función que permita preparar datos de entrada para su modelo entrenado y una función que ayude a interpretar la salida de su modelo.