



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Universidad de Buenos Aires Facultad de Ciencias Económicas Escuela de Estudios de Posgrado

CARRERA DE ESPECIALIZACIÓN EN MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE DATOS EN ORGANIZACIONES

TRABAJO FINAL DE ESPECIALIZACIÓN

Utilización de machine learning para la predicción del
comportamiento transaccional de clientes de compañías
Fintech con el fin de detectar anomalías

AUTOR: NICOLÁS MANA

DICIEMBRE 2019

Resumen

En el siguiente trabajo se estará desarrollando la problemática de la empresa “Fintech Argentina S.A.”, compañía de procesamiento de pagos por internet, la cual quiere mejorar la calidad de servicio de soporte de sus clientes. Para lograr esto, se plantea desarrollar un sistema que pueda predecir la cantidad de transacciones que reciben sus clientes, y así disparar alertas ante un escenario que presente anomalías en dicho comportamiento, y poder accionar sobre el mismo de manera proactiva. Se considera que tomando el histórico transaccional de sus clientes se puede predecir el comportamiento futuro mediante la utilización de técnicas de Machine Learning, y utilizar dicha predicción como base para el disparador de alertas.

Se estará desarrollando el contexto en el cual se encuentra la organización, el detalle de su problemática y la metodología a utilizar para poder obtener la mencionada predicción.

Palabras clave: Fintech, Machine Learning, Predicción transaccional, Calidad de servicio

Estructura

Introducción	4
Fintechs y Big Data.....	6
Historia y escenario actual	8
Relación con Big Data	10
La organización y su problemática	11
Machine Learning en la práctica.....	14
Base de datos.....	14
Elección e implementación del modelo	16
Conclusión del modelo	20
Impacto organizacional del modelo	21
Implementación y mejora del modelo en la organización	21
Oportunidades de mejora	22
Propuestas futuras	22
Conclusión	23
Referencias bibliográficas.....	25
Anexos	26
Apéndices.....	26

Introducción

Desde hace ya varios años, las compañías que aplican nuevas tecnologías para mejorar las actividades financieras, denominadas Fintechs, han tenido un crecimiento exponencial en todo el mundo, incluyendo Argentina. Se ha registrado un aumento significativo en el uso de teléfonos celulares y de Internet para realizar operaciones financieras. Entre 2014 y 2017, esto ha contribuido a que aumente del 67 % al 76 % a nivel mundial, y del 57 % al 70 % en el mundo en desarrollo, la proporción de titulares de cuentas que envían o reciben pagos a través de medios digitales. A nivel mundial, si bien hay 1700 millones de adultos que aún no están bancarizados, dos tercios de ellos poseen un teléfono celular que podría ayudarlos a acceder a los servicios financieros. En América Latina y el Caribe, el acceso amplio a la tecnología digital podría permitir el crecimiento acelerado del uso de tecnología financiera: el 55 % de los adultos tiene teléfono celular y acceso a Internet, es decir, 15 puntos porcentuales más que el promedio del mundo en desarrollo. Desde 2014, la proporción de adultos que realiza o recibe pagos digitales ha aumentado unos 8 puntos porcentuales o más en economías como las de Bolivia, Brasil, Colombia, Haití y Perú. En Argentina, Brasil y Costa Rica, alrededor del 20 % de los adultos que tienen cuenta usa el celular o Internet para operar a través de una cuenta. Digitalizando los pagos de sueldos en efectivo, las empresas podrían contribuir a que hasta 30 millones de adultos no bancarizados —de los cuales casi el 90 % tiene teléfono celular— sean titulares de una cuenta. (Asli Demirgüç-Kunt, 2017). Particularmente en Argentina, compañías como NuBank, TodoPago, Ualá, Mercado Pago, entre otras, han podido insertarse en el mercado financiero Argentino ofreciendo nuevas soluciones tanto para el mercado bancarizado como el no.

Teniendo en cuenta el potencial crecimiento que tienen estas empresas y el valor primordial que tienen los datos en este tipo de organizaciones, cuya premisa es la utilización de la tecnología para mejorar la calidad de servicio financiero, se estará abordando la problemática de cómo utilizar la gran cantidad de datos que poseen para poder aportar el mayor valor agregado posible al servicio que le brindan a sus clientes.

Dentro de la categoría Fintech podemos encontrar distintos tipos. Encontramos empresas orientadas a manejo de inversiones personales, crowdfunding, prestamos, seguros y, sobre la cual nos estaremos centrando, procesamiento de pagos por internet. Nuestro caso de estudio se basará en el tratamiento de la información obtenida para la empresa “Fintech Argentina S.A.”

En Argentina, este tipo de empresas procesan tanto pagos por internet de clientes muy pequeños, que pueden llegar a procesar mil dólares por mes, como clientes muy grandes que pueden llegar a procesar millones de dólares en el mismo período de tiempo. Dentro de este segundo grupo, se genera un subgrupo conocido como la “cartera atendida”, que hace referencia a los principales vendedores, es decir, quienes procesan la mayor cantidad de pagos por internet, y son los cuales cuentan con una atención premium, con soporte técnico y comercial más personalizado. Debido a que cada vez más el servicio de procesamiento de pagos por internet se está volviendo un commodity en Latinoamérica y el mundo, la calidad de servicio que se le pueda dar a los clientes es el diferencial para poder obtener adopción y ganancias del mismo.

Para brindar una atención de calidad superadora a los clientes de la cartera atendida, se debe responder de manera inmediata ante cualquier tipo de inconveniente que pueda tener el procesamiento de un cliente en particular. Para lograr un accionar lo más rápido posible frente a estos problemas, ¿como se puede, analizando el comportamiento transaccional de la cartera asesorada, predecir el mismo y detectar anomalías?

Con el fin de poder responder esta incógnita se va a tomar el registro de las últimas transacciones que tuvieron los 10 principales clientes de dicha cartera en los últimos 4 meses, y se va a utilizar un modelo de machine learning para predecir que cantidad de transacciones deberían recibir los mismos en las horas/día siguiente. Se utilizará la predicción como disparador de alertas en los casos que difiera mucho el incoming de las transacciones con respecto a la predicción obtenida. El objetivo principal del trabajo se desglosará en base a este accionar de la siguiente manera. Se obtendrá el histórico de los clientes (el cual tendrá un tratamiento previo a su análisis) y luego se utilizará el mismo para entrenar un modelo predictivo que entregue las bases para las futuras alertas, mejorando la calidad de servicio de “Fintech Argentina S.A.”

Para considerar un cliente parte de este grupo, tuvo que haber tenido operaciones por un mínimo de 3 meses para que los equipos de soporte puedan brindar sus servicios, por lo que se va a tomar en el análisis los primeros 3, sumado a un cuarto para favorecer al entrenamiento del modelo. La idea es partir de esta base de 4 meses para después seguir sumando información de los períodos de tiempos futuros con el fin de seguir entrenando el modelo predictivo. Si bien el modelo es extrapolable a cualquier empresa Fintech de Latinoamérica, se va a estar analizando particularmente el caso de Argentina, teniendo en cuenta variables locales que puedan afectar el comportamiento transaccional de los clientes,

como si es un fin de semana, las fechas son cerca de alguna festividad, eventos como el “Cyber Monday”¹, feriado, etc.

En los siguientes tres apartados se estará desarrollando con mayor detalle el trabajo. Comenzará con la explicación de cómo la empresa “Fintech Argentina S.A.” está abordando esta problemática hoy en día, cual es el contexto en el cual se encuentra y porque necesita un cambio de enfoque con respecto a su forma de trabajo y tratamiento de datos. Luego se detallarán los aspectos metodológicos del trabajo, como se consiguieron los datos, como se trataron, que tipo de modelo predictivo se utilizó y las implicancias más técnicas referidas a estos puntos. Y finalmente se abordará la propuesta para la implementación del trabajo final y los beneficios que este conlleva.

Fintechs y Big Data

En este apartado se estará desarrollando el estado actual del conocimiento. Se detallará sobre las Fintech, su historia y los desafíos de las mismas. Además, se estará detallando la relación del mundo de las Fintech con la problemática organizacional con la que cuenta “Fintech Argentina S.A.” y las bases para su futuro abordaje en los apartados posteriores.

Para comenzar a desarrollar el trabajo y tener la base de conocimiento para poder ir abordando las problemáticas planteadas y los métodos utilizados, primero tenemos que definir ciertos conceptos.

Fintech

El término fintech deriva de los términos financial y technology. Financial, porque estas empresas brindan servicios específicos en los sistemas financieros. Technology, porque estos servicios funcionan con tecnologías que generalmente son novedosas. (Velarde, 2019)

Su propuesta está centrada en algún aspecto concreto de las finanzas (préstamos, capacitación de recursos, medios de pago, análisis de datos, asesoramiento financiero, etc). Esta simplicidad basada en el concepto de una única propuesta de negocio muy vertical y centrada en un área específica, contrasta con los actuales complejos servicios de los bancos con prestaciones que son difíciles de encontrar y ejecutar en sus webs. Proceden de la cultura de la innovación y de las startups creadas desde cero, con una filosofía de romper con los

¹ Cyber Monday: Tres días de descuentos en compras online impulsadas por la Cámara Argentina de Comercio Electrónico en todo el país.

anteriores formatos. Las nuevas tendencias de economía colaborativa lazan directamente con la filosofía de estas nuevas empresas, y por ello conectan más fácilmente con el perfil de los nuevos consumidores. En este contexto, la propia opinión de los usuarios sobre los servicios y la experiencia del cliente en su consumo es más apreciada por estos que ser un cliente de un gran banco con una marca muy poderosa pero que inspira poca confianza. (Iguar, 2016)

Las Fintech abarcan un conjunto de empresas muy distintas que, atendiendo a su naturaleza, se pueden clasificar en tres grupos. “*Startups y compañías de pequeño tamaño*” es el primer grupo que representa a la pequeña o mediana empresa de creación reciente y, normalmente, vinculada con el mundo de la tecnología. Su objetivo principal es instrumentar su producto y escalar su solución hasta alcanzar una suficiente masa crítica que les permita una autonomía financiera. Luego se pueden encontrar a las empresas “*Unicornios*”, que son startups cuya valoración es superior a los mil millones de dólares. El modelo de negocio de mayor éxito obedece a tener una amplia base de clientes que ofrecen un elevado potencial de obtención de ingresos y su actividad se plantea en términos de independencia con respecto a los bancos. Y finalmente las “*GAFAS*” (Google, Amazon, Facebook y Apple), también conocidas como bigtechs, el grupo hace referencia a las corporaciones de mayor capitalización del mundo. Se considera el grupo debido al gran poder intimidatorio que significan para la banca tradicional por el potencial que tienen de penetrar el negocio financiero y desplazar a los bancos. (Iguar, 2016)

A su vez, se presenta una clasificación basada en los tipos de actividades que realizan las mismas, en la que se encuentran algunos de los nuevos servicios financieros que brindan estas empresas en forma particular. Esta clasificación tiene seis grandes tipos. “*Pagos y transferencias*”, empresas que ofrecen soluciones sobre medios de pago electrónicos, canales digitales y transferencias que permiten la ejecución de transacciones entre los diferentes agentes económicos, evitando el empleo del dinero físico y de forma independiente respecto a los bancos. “*Préstamos y financiamiento*”, espacio electrónico que pone en contacto a los inversores con los solicitantes de financiación. “*Banca minorista*”, plataformas dedicadas a la inversión y gestión del patrimonio para particulares y pequeñas y medianas empresas. “*Gestión financiera*”, plataformas que permiten a sus usuarios ver a través de dispositivos móviles sus cuentas, movimientos, facturas, etc. de forma segura, agregada y simple. “*Seguros*”, empresas que han desarrollado innovaciones en temas de seguridad como la identificación y la gestión de la identidad digital. Y finalmente “*Mercados y financiación*”, plataformas en donde los inversores a proyectos identificados y validados, desde cantidades pequeñas. (Devis, 2018)

Predicción y análisis de anomalías

Cuando se habla de anomalías, se hace alusión a un cambio o desviación de lo que es normal o previsible. En el desarrollo de este trabajo se habla de anomalías a la hora de analizar la cantidad de transacciones que tiene un cliente a lo largo del tiempo y su desviación correspondiente. La utilización de modelos de machine learning para predicción del comportamiento transaccional y la detección de anomalías con el fin de detectar fraudes o problemas en el servicio puede ser muy variado. El crecimiento del ecommerce es acompañado por un crecimiento de los niveles de fraude. Investigadores han propuesto reglas basadas en sistemas de auditoria para las transacciones online, pero dependen mucho del conocimiento del auditor sobre el fraude en el ecommerce, por lo que se ve afectado. Mientras los patrones de fraude pueden generarse, los controles de dichos patrones es difícil debido a la gran cantidad de transacciones que se manejan hoy en día. (Monkol Lek, 2001). Por estas razones un modelo de machine learning favorece a que estos análisis sean lo más objetivos posibles.

Es imposible estar seguros de la legitimidad o intención detrás de una aplicación o transacción. Dada la realidad, la opción más efectiva es aproximar ciertas evidencias de fraude/anomalías desde los datos disponibles utilizando algoritmos. (CLIFTON PHUA1*, 2010)

Historia y escenario actual

La industria fintech presenta un importante desarrollo en los últimos años, principalmente en Norteamérica, Europa y Asia, el que es impulsado por diversos aspectos económicos, legales y tecnológicos (Lee, 2018).

Las primeras Fintech, aunque sin esta denominación, tienen su origen en el 2008, aunque su impulso y desarrollo notable no se produce hasta el 2010. El predominio de estas empresas hoy en día con origen en Estados Unidos y China es muy notable, y la inferior presencia de fintechs europeas obedece a la menor dimensión de estos mercados, así como la diversidad de idiomas y culturas existentes en los diferentes países europeos, que las convierten en mas locales y reducen la potencialidad. (Igual, 2016)

Desde comienzos de 2010, más de 24 billones de dólares han sido invertidos en startups orientadas a Fintech en todo el mundo. Más de 11 billones de ese total fueron invertidos en la primera mitad de 2015. ¿Que puede explicar el boom en inversión de ese tipo de empresas? La crisis de créditos de 2008 y el alto nivel de regulaciones sobre los servicios

financieros que siguieron crearon un ambiente perfecto para la innovación en la banca. Siguiendo la crisis financiera, los bancos se encontraron con nuevas regulaciones y penalidades. Estas restricciones obligaron a los bancos a prestar más atención a sus procesos internos y gastar más dinero en gobernanza de datos y manejo del riesgo como nunca antes. Las startups Fintech pudieron crear modelos de negocio que evitaron las estructuras formales de un banco mientras ofrecían servicios mas adecuados a las necesidades de los clientes.

El alto nivel de regulaciones alrededor de los servicios financieros intenta ser el mismo a nivel global por una buena razón. Los servicios financieros manejan los activos de las personas, ahorros para jubilaciones y salarios. Las regulaciones protegen a los clientes e inversores teniendo el sistema en control. Debido a esto, los fundadores de las Fintech tienden a ser ex empleados de bancos que brindan conocimiento y entendimiento sobre los procesos y problemas principales que deben ser solucionados. Pero no todo es perfecto en el mundo de las Fintech, generalmente tienen inconvenientes a la hora de querer expandirse. La gran cantidad y complejidad de regulaciones para operar en cada país hacen que las operaciones que crucen fronteras de los países sean legalmente costosas y riesgosas para las compañías, uno de los principales problemas en los países emergentes. (Desai, 2015). A su vez, tienen expectativas de desarrollo porque cuentan con mayor margen y oportunidades para la innovación en servicios financieros, en comparación con los países que tienen economías desarrolladas, exceso de entidades bancarias y sistemas financieros maduros y estables (Chishti, 2016).

En América Latina las startups financieras están recibiendo inversiones significativas de diferentes orígenes, incluyendo la banca tradicional, llamada Finvestment. Algunos startups tienden a convertirse en bancos hacia delante. Por ejemplo, Nubank en Brasil, un operador de tarjeta de crédito, ofrece tarjetas de crédito mpbiles a 4 millones de usuarios sin ningún tipo de tarifa, llegando a ahorros digitales de usuarios de 1.5 millones de dólares para Junio 2018. Al mismo tiempo, otros Startups ofrecen abrir una caja de ahorro accesible a través del teléfono celular y una tarjeta de débito que puede ser utilizada físicamente. En la región, las empresas se centran principalmente en pagos y alternativas financieras. (P. Berkmen, 2019). Debido a este crecimiento estructural y a su desarrollo en productos y servicios, “Fintech Argentina S.A.” ha tenido un crecimiento muy importante en los últimos años acompañando a la región, duplicando la cantidad de transacciones que procesa trimestre a trimestre desde principios de 2018.

Relación con Big Data

Big data es uno de los aspectos claves de la transformación digital que afecta a todos los sectores de la actividad como consecuencia del incremento inmenso de datos que se generan, susceptibles de ser aprovechados de forma eficiente. Hace referencia al tratamiento y análisis de bases de datos lo suficientemente grandes como para ser procesadas por sistemas tradicionales, con la finalidad de obtener percepciones que tengan impacto en el negocio. Diferentes empresas de investigación tecnológica afirman que el 90% de los datos que existen se han creado en los últimos años, e incluso que el 75% de los datos actuales no son estructurados y provienen de fuentes tales como la voz, texto o video. El big data en sí no sirve de nada si no hay una estrategia atrás que plantee las grandes preguntas que se espera que los grandes volúmenes de datos ayuden a responder. Por ejemplo, desarrollar una estrategia de personalización de la oferta, detección de fraude o predicciones de consumo. (Igual, 2016)

La utilización de herramientas de machine learning son primordiales para el desarrollo de empresas en contextos disruptivos como lo puede ser Latinoamérica. El aprovechamiento de la información para obtener un diferencial sobre la competencia puede ser un factor clave con respecto al crecimiento de la misma.

En 2012 el MIT Sloan Management review junto a IBM hicieron un estudio sobre 3000 ejecutivos de alto nivel de 100 empresas a lo largo de 30 industrias para analizar el impacto de la utilización de Data Analytics sobre el rendimiento de la misma. Data Analytics hace referencia a procesar y analizar datos para descubrir como los mismos pueden ser utilizados para conclusiones o resolución de problemas de negocio.

Las organizaciones que coincidieron que el uso de la información de negocio y business analytics las diferencian de la industria, tienden a ser las mejores de su categoría. Los “Top performers” de cada categoría abordan los problemas de negocio de una manera distinta, tienden a utilizar herramientas de data analytics para cualquier tipo de decisión, sea grande o pequeña. A pesar de la opinión popular de que obtener los datos es la parte más difícil que las organizaciones afrontan a la hora de adoptar técnicas de analytics, solamente una de cinco respuestas coinciden en este como un principal problema. Una de las principales preocupaciones en general es las barreras de adopción hacia estas nuevas tecnologías. El principal obstáculo es vender la solución lo suficientemente bien para poder obtener más fondos y recursos que favorezcan a la obtención de resultados en el corto plazo. (LaValle, 2012)

Empresas de varias industrias tradicionales están explotando nuevos y existentes recursos de datos con la finalidad de obtener una ventaja competitiva. Contratan equipos de

científicos de datos que puedan utilizar nuevas tecnologías para aumentar el retorno y disminuir los costos. Además, las nuevas compañías están siendo desarrolladas con minería de datos como un componente principal de su funcionamiento. Facebook y Twitter, junto con otras compañías del “Digital 100”² tienen un alto valor debido a la gran cantidad de datos que poseen o tienen la posibilidad de crear. (Fawcett, 2013) Los coporativos de alto nivel en las empresas deben manejar conceptos de análisis de datos para poder aprovechar al máximo la nueva “era de los datos”.

Con respecto a la producción de servicios financieros, una gran cantidad de información personal es recolectada y utilizada hoy en día. Grandes bases de datos de clientes y el histórico de su comportamiento permite predecir preferencias, necesidades y tendencias, y ofrecer el producto financiero correcto, en el momento correcto al precio correcto. Big data y machine Learning son dos ingredientes principales que están cambiando la industria de servicios financieros. Las aplicaciones y los efectos en la industria están creciendo y lo seguirán haciendo, desde análisis de potenciales clientes hasta el análisis de riesgo a nivel individual de una empresa. (Vives, 2017)

La organización y su problemática

Como se desarrolló previamente, “Fintech Argentina S.A.” es una empresa de tecnología que brinda servicios financieros utilizando innovaciones tecnológicas. Particularmente se encuentra en la categoría de “Pagos y transferencias”, ya que procesa los pagos por internet de cada uno de sus clientes.

Los equipos comerciales que tienen relación con los clientes que utilizan los servicios de una empresa se dividen en dos, equipos de “hunting”, que son los que salen a buscar nuevos clientes, y los equipos de “farming”, los cuales se encargan de brindarle soporte a los clientes actuales y aprovechar la relación para poder realizar cross y up-selling³. Muchos clientes contratan los servicios por el buen trabajo del equipo de hunting, pero el buen desempeño del equipo de farming asegura que los clientes mantengan el servicio generando ganancias y confianza a largo plazo. Un mal proceso de farming puede generar la pérdida de clientes muy valiosos y mala reputación que ahuyentaría a potenciales clientes

² Digital 100: Fusfeld A. The digital 100: the world’s most valuable startups. Bus Insider. Sep. 23, 2010.

³ Up selling: Se le vende aún más a un cliente ya activo. Cross selling: Se le vende un producto nuevo a un cliente activo.

nuevos. Una vez que el cliente es captado por el equipo de hunting y pasa 3 meses procesando con la empresa, el soporte del mismo para estar a cargo del equipo de farming.

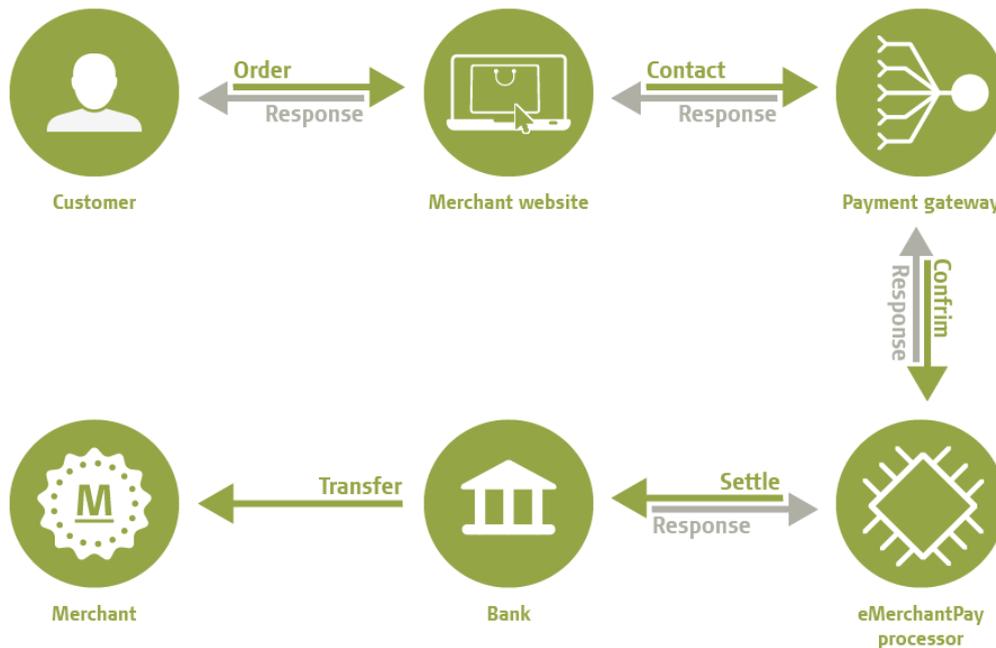
No hay un proceso formal para la inducción de un cliente nuevo al proceso de soporte en “Fintech Argentina S.A.”, sino que simplemente se van atendiendo clientes a medida que se van presentando problemas, los cuales se toman con una determinada prioridad dependiendo del tipo de cliente (small, médium, large) y de la proyección a futuro que tiene el mismo. Si bien es una mejora al proceso que se implementaba antes, en el cual se tomaban a todos los clientes por igual, está muy lejos de ser un proceso ideal. Actualmente se está intentando reestructurar este proceso, en el cual se hará un traspaso mucho más claro de cara al cliente de un equipo a otro, indicando mejoras y canales de soporte principales para cada cliente.

Según un estudio realizado por Mckinsey en septiembre de 2019 (Company, 2019), los tres principales factores que van a diferenciar las empresas de procesamientos de pago en los próximos años serán las propuestas basadas en el cliente, por ejemplo, pagos crossborder⁴, capacidades omnicanal y experiencia de usuario distintiva, como facilidad para la integración y simplicidad de la solución. Escala, por ejemplo, costos fijos consolidados y productos que se adapten a nuevos negocios. Y el tercer factor diferenciar estará relacionado a las “Capabilities”, por ejemplo, tecnología para procesamiento de pagos, automatización de procesos, respuesta ante contingencias y acceso en tiempo real a métricas y reportes de conciliación junto con analytics que estén orientados a entregar valor al cliente.

Teniendo en cuenta estos puntos, una de las mejoras que se busca implementar en “Fintech Argentina S.A.” es poder brindar soporte de calidad a los clientes y de manera proactiva, detectar fallas en el sistema antes de que afecte el negocio del cliente y sea demasiado tarde. Un sistema predictivo que detecte anomalías en el funcionamiento del servicio del cliente es el eslabón clave para que el nuevo proceso de traspaso al equipo de farming funcione de la mejor manera, generando un diferencial con la competencia.

Para brindar el servicio correspondiente, “Fintech Argentina S.A.” tiene que comunicarse con varios interlocutores, lo que genera más puntos de riesgo para mantener la calidad del servicio. Tiene un modelo que le permite operar como si fuera únicamente un canal de conexión entre el adquirente de la tarjeta (Visa, Master, Amex, Etc) y el cliente. A este tipo de procesamiento de pagos se lo conoce como Gateway de pagos.

⁴ Crossborder: Pagos entre usuarios de distintos países



El comprador final ingresa los datos de tarjeta de crédito en el sitio web del vendedor, “Fintech Argentina S.A.” luego se comunica con el adquirente de la tarjeta (Visa, Master, Amex, Etc), y el mismo con el banco para ver si la tarjeta tiene fondos y puede aceptar el pago. Hay varios puntos del flujo que pueden llegar a tener un problema, lo que afectaría el servicio que “Fintech Argentina S.A.” le brinda al vendedor final. El fin del trabajo no es mitigar estos riesgos, sino poder accionar lo más rápido posible ante la ocurrencia de algún problema relacionado al mismo.

Teniendo estos conceptos en claro, podemos ver que la explotación de los datos con los que cuenta “Fintech Argentina S.A.” son primordiales para el crecimiento del negocio tanto personal como del de la industria. Una de las principales diferencias que puede ofrecer una empresa Fintech es el soporte que le brinda a sus usuarios, tratando de accionar ante cualquier problema que pueda ocurrir en el flujo de pago de manera proactiva. Utilizar la información que tiene disponible para detectar mejoras y predecir necesidades de los clientes es lo que puede darle el diferencial a un prospecto a la hora de elegir la empresa como proveedora de servicio o no.

En el próximo apartado se estará detallando cómo se hará para aprovechar al máximo la información con la que cuenta “Fintech Argentina S.A.”, definiendo la elección y aplicación de herramientas y modelos predictivos.

Machine Learning en la práctica

En el apartado anterior se pudo ver el estado actual de conocimiento sobre la temática abordada en el trabajo, definiendo ciertos conceptos previos y entrando en detalle en la histórica y situación actual de las problemáticas que afectan a “Fintech Argentina S.A.” con respecto al manejo de grandes volúmenes de datos.

En este segundo apartado se va a estar desarrollando la parte más práctica del objetivo del trabajo, que es poder predecir el comportamiento transaccional de los principales clientes de la cartera asesorada de “Fintech Argentina S.A.” para poder detectar anomalías de manera proactiva. Dicha predicción se realizará utilizando modelos predictivos de machine learning, teniendo como base de entrenamiento el histórico de las transacciones procesadas por cada cliente a lo largo de los últimos cuatro meses. En sí primero se obtendrá el histórico de los clientes. Se obtendrá el histórico consultando los recursos que presenta la empresa y se hará una depuración de los mismos. Luego se terminará el procesamiento de los datos. Una vez se tengan los datos, se utilizarán herramientas como Excel y SQL para el almacenamiento y tratamiento de los mismos, así como también se definirán las herramientas a utilizar para aplicar el modelo predictivo elegido. Y finalmente se va a predecir en base al histórico. Una vez se tengan los datos necesarios para realizar la predicción, la misma se realizará con un modelo predictivo de machine learning para obtener el porcentaje de precisión.

Base de datos

“Fintech Argentina S.A.” cuenta con el histórico de todas las transacciones de sus clientes desde que comenzaron a operar con ellos hasta la fecha de la última transacción. Utilizando los criterios y técnicas de “Gestión de datos en contextos organizacionales”, para mantener la integridad y seguridad de los datos que proporcionó la compañía para el trabajo, todos los datos fueron anonimizados para poder trabajar sobre los mismos sin comprometer la seguridad de los clientes de la compañía.

Se comenzó tomando toda la base de grandes clientes de Argentina, Chile y Uruguay desde comienzo del año, pero utilizando dichos parámetros se encontraron dos principales problemas. El primero es que los contextos de industria y país en los tres países son muy diferentes a lo largo del año, afectando las predicciones de las empresas (o empresas dentro de una misma industria) que operan en los tres países, por lo que para un primer acercamiento se tomaron como mercados/bases de predicción separadas. Y segundo, los clientes no son los

mismos a lo largo del año, muchos comienzan y dejan de operar con “Fintech Argentina S.A.” mientras transcurren los meses, sumado a que para que formen parte de la “cartera atendida” tienen como requisito haber tenido mínimamente tres meses de transacciones procesadas con la compañía. Para solucionar estos inconvenientes se trabajó principalmente con el mercado Argentino, sobre el cual “Fintech Argentina S.A.” tiene mayor penetración, y solamente se tomaron los principales diez clientes que hayan tenido transacciones en los últimos 4 meses.

Internamente el histórico de las transacciones se guardan en bases de datos alojadas en la nube pudiendo ser consultada de dos maneras distintas. La primera es mediante MySQLServer con acceso a los servidores que guardan las bases de datos, y la segunda es mediante una API Rest, con accesos restringidos, que retorna la información en formato .csv. Ambos métodos devuelven la cantidad de transacciones de un cliente en el período de tiempo que se le indique incluyendo el “id del usuario”, “fecha” (día/mes/año hora), “industria” del cliente y “ticket promedio” (el valor promedio que tienen las transacciones que procesa cada cliente).

Por cuestiones prácticas, se utilizó la respuesta de la API Rest tomando los últimos cuatro meses y catorce días de los principales diez clientes de la cartera asesorada desde el 14 de septiembre de 2019 hacia atrás. Es decir, el histórico desde el 1/5/2019 hasta el 14/9/2019 para intentar predecir, en un principio, el comportamiento transaccional de cada cliente durante el 15/9/2019. En una implementación productiva la base se iría actualizando día a día teniendo un registro más amplio y prediciendo el día siguiente con mayor precisión a medida que se entrena el modelo. Una de las razones por las cuales se tomó ese rango de tiempo, es porque los principales 10 clientes de la cartera atendida tienen operaciones en dichas fechas, teniendo una base lo suficientemente llegando a 36.000 registros.

Con el fin de aprovechar al máximo la información que provee el csv brindado por “Fintech Argentina S.A.”, se realizó un tratamiento de la base que consiste en dos frentes principales. El primero fue separar el campo “fecha” en “día”, “mes”, “año” y “hora”. Y el segundo en agregar la mayor cantidad posible de campos a la base para aportar información que pueda ayudar a la predicción, entre ellos se agregaron “Día de la semana” (Valores: 1-Lunes, 2-Martes, 3-Miercoles, 4-Jueves, 5-Viernes, 6-Sabado y 7-Domingo), “Es fin de semana?” (Valores: 0-No y 1-Si), “Es feriado?” (Valores: 0-No y 1-Si), “Está en algún evento?” (Valores: 0-No y 1-Si) y “Cuántos días faltan para el próximo evento?” (Valores: 0-365).

Con “evento” se refiere a una fecha importante a lo largo del año que pueda afectar la cantidad de transacciones que se realizan en internet debido a la gran cantidad de descuentos y oferta de cuotas sin interés que se promocionan. Para las fechas que se tomaron, los eventos más importante son “Hot Sale” (14-16/5), “Día del padre” (16/6) y “Día del niño” (18/8). Se creó una base aparte con los eventos más importantes del ecommerce durante el año llamada “eventos” para poder actualizar las futuras bases del proyecto, sabiendo si es o no un evento importante y cuantos días faltan para el próximo, así como también si fue feriado o no.

Para una mejora en el tratamiento y modificación de las futuras bases, se utilizará un script en Python (prepare_post_create.py) para poder realizar todas las modificaciones previamente mencionadas de manera automática con el fin de llegar a la base que efectivamente se estará utilizando como input del modelo predictivo, así como también la base del día a predecir.

En resumen, utilizando la respuesta de la API, la base de eventos y un script para su correcta adaptación, se llegó a la base que se utilizará para entrenar al modelo predictivo, la cual llamaremos “tp_modelo_entrenamiento”, y la base con la información necesaria para obtener la predicción del día siguiente, la cual llamaremos “base_a_predecir”. El detalle de los recursos mencionados (“Respuesta API”, “eventos”, script, “tp_modelo” y “base_a_predecir”) se encuentran en Apéndices, así como también se adjuntan al trabajo como soporte al mismo.

Elección e implementación del modelo

En esta sección se detalla un poco sobre como se eligió el modelo predictivo a utilizar y el proceso de implementación del mismo. Una vez obtenida la base “tp_modelo_entrenamiento” se tuvo que elegir el modelo predictivo a utilizar para realizar la predicción y la herramienta para poder aplicar el mismo.

La plataforma elegida para realizar los análisis fue Microsoft Azure Machine Learning Studio. Es un servicio de aprendizaje automático de nivel empresarial para crear e implementar modelos de Machine Learning con más rapidez. Aumenta la capacidad de los desarrolladores y científicos de datos con una amplia gama de experiencias productivas para crear, entrenar e implementar modelos de Machine Learning en la nube con mayor velocidad. También se tuvieron en cuenta las herramientas de RapidMiner y GoogleColab Notebook (para desarrollo en Python), pero debido a las facilidades de comparación/ejecución de modelos predictivos con una interfaz de usuario amigable, la posibilidad de definir un web service mantenido por la

infraestructura de Microsoft y la capacidad de procesamiento se decidió utilizar Azure Machine Learning.

Dentro de la herramienta, se utilizó un experimentos desarrollado por Microsoft para poder comparar los modelos predictivos. Se compararon:

- “Bayesian linear regression”: Se utiliza regresión lineal suplementada con información adicional de distribución de probabilidad
- “Neural Network Regression”: Se utiliza un modelo de regresión lineal utilizando algoritmos de redes neuronales.
- “Boosted Decision Tree Regression”: Utiliza una regresión basada en el algoritmo de “Boosted decisión Tree”, el cual crea árboles de decisión midiendo el error en cada paso del árbol para corregirlo en el siguiente eligiendo el óptimo al final del proceso.
- “Linear Regression”: Utiliza un modelo de regresión lineal.
- “Decision Forest Regression”: Utiliza un modelo de regresión con el algoritmo de random forest.

Luego de comparar los modelos con la base de “tp_modelo_entrenamiento”, en donde se tomó una una parte de la base como entrenamiento del modelo y otra para medir la exactitud de la predicción utilizando “Cross validation” y evitar overfitting⁶, se llegó a la conclusión de que el mejor modelo a utilizar para este caso teniendo en cuenta el error cuadrático medio de la predicción fue el “Boosted Decision Tree Regression”. Se tomó dicha medida de evaluación ya que ayuda más fácilmente a comparar si las transacciones reales que se vayan midiendo se encuentran lo suficientemente cerca (dentro del rango definido por el error cuadrático medio) de la predicción. En base a eso se definiría el aviso de las alarmas.

⁵ Experimento: Nombre de los proyectos en Azure Machine Learning Studio

⁶ Overfitting: Sobredeterminación o sobreaprendizaje. Si el ajuste es muy fuerte el modelo deja de ser generalizable y sólo refleja las particularidades del conjunto utilizado en la prueba.

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determinatio
Bayesian Linear Regression	14.166452	27.369423	0.856796	0.807638	0.192362
Neural Network Regression	11.012466	20.937052	0.666041	0.472625	0.527375
Boosted Decision Tree Regression	3.935322	7.862879	0.238011	0.066657	0.933343
Linear Regression	14.166893	27.369187	0.856822	0.807624	0.192376
Decision Forest Regression	3.864406	8.967365	0.233722	0.086699	0.913301

Con la decisión tomada de que modelo se utilizará, se procedió a utilizar la base “tp_modelo_entrenamiento” como input de entrenamiento de “Boosted Decision Tree Regression” para intentar predecir la cantidad de transacciones de cada cliente por horario para la base “base_a_predecir”. Para ello, se creó el experimento predictivo “TP-MANA-Prediction”, en el cual con el modelo de “Boosted Decision Tree” entrenado se utilizó la “base_a_predecir” para obtener las predicciones del 15/9. Los parámetros de dicho experimento se definieron al realizar pruebas con un “Split data”⁷ de la base en un 80/20 para aplicar el modelo predictivo al 80% de la misma y utilizar la otra para comparar los resultados obteniendo distintos grados de precisión.

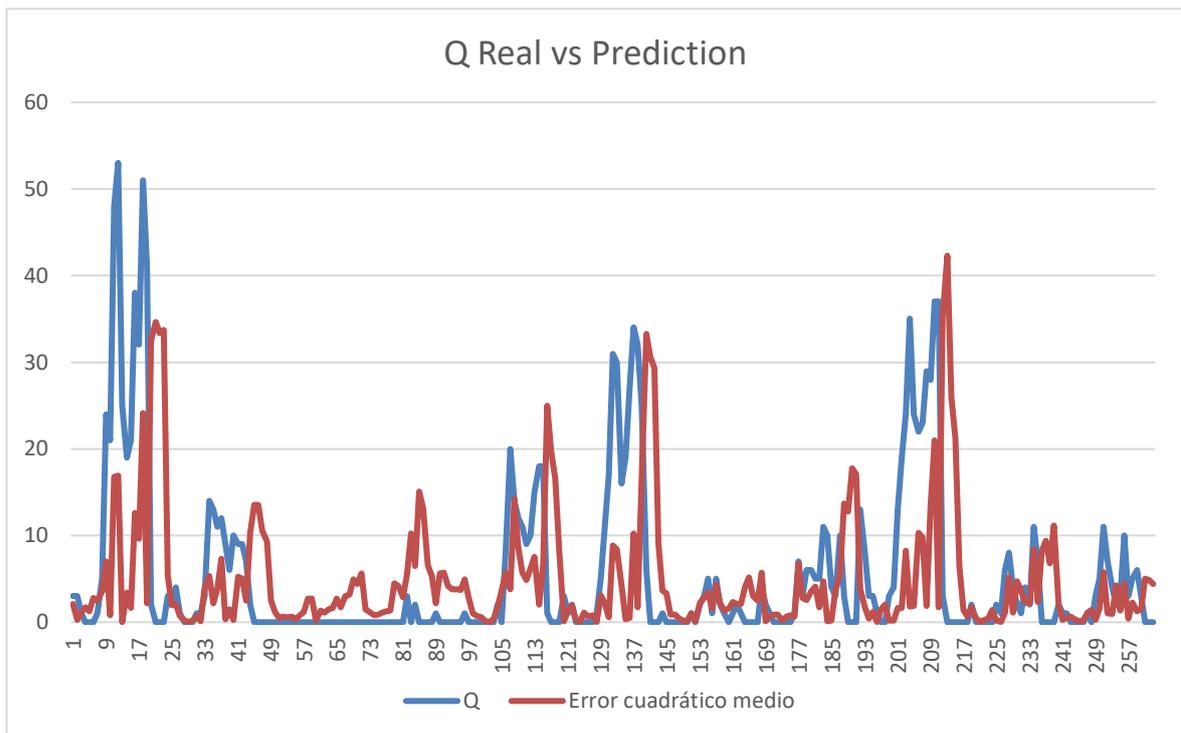
Para la futura escalabilidad del proyecto, el experimento se publicó como un web service predictivo. El servicio de API de Azure Machine Learning permite publicar modelos predictivos armados en Azure como Web Services escalables y tolerantes a fallos. Se recibe un base en formato JSON a predecir, se ejecuta el modelo en la nube y se devuelven los resultados en la respuesta de la API call hacia el servicio.

Mediante la utilización del script “prepare_post_create.py” se convirtió la “base_a_predecir” a un formato JSON y mediante un POST se llamó al webservice que retornó luego de 24 segundos la base con la predicción, la cual se guardó en un archivo .json

⁷ Split data: Se separa la base.

y en un .xlsx para su futura utilización en las alarmas (prediction_fecha.json y prediction_fecha.xlsx)

Una vez obtenida la predicción para el 15/9 se comparó con el comportamiento transaccional real de los clientes de la cartera asesorada para esa fecha obteniendo una predicción con un error cuadrático medio de 5 (menor a la media que se había obtenido para la base “tp_modelo” de 7,8).



El objetivo a futuro es que a medida que se vayan sumando más registros a la base predictiva “tp_modelo_entrenamiento” se pueda seguir entrenando el modelo para que el error medio siga disminuyendo. A su vez, se ingresará como parámetro si la predicción previa fue exitosa o no para ayudar al ajuste del modelo a futuro.

En caso de que se quieran predecir más días juntos (no solamente el 15/9) se tiene que crear un jobs en Azure Machine Learning Studio para darle más tiempo de procesamiento, ya que el método de request/response de la API no soporta procesos que duren más de 90 segundos.

8 Job: Proceso que se ejecuta automáticamente sin necesidad de interacción del usuario.

Conclusión del modelo

Para poder ejecutar el modelo predictivo previamente entrenado se siguió el siguiente procedimiento:



Se pudo obtener un modelo que cumpla el objetivo principal del trabajo para predecir el comportamiento transaccional de los clientes. El modelo de “Boosted Decision Tree Regression” terminó siendo el que mejor se adaptó a las necesidades de este proyecto. Se le definieron 200 hojas por árbol de decisión con un total de 500 árboles que se contruyen para la definición del modelo. Se presentó una mejora considerable con respecto a la utilización de los otros modelos llegando a un accuracy más grande cuando se lo compara con el rendimiento real de los clientes durante el 15/9.

En el próximo apartado se va a estar abordando la propuesta del trabajo, el desarrollo del mismo, como se relaciona con el objetivo general y la evaluación de la solución presentada.

Impacto organizacional del modelo

En el apartado anterior se pudieron presentar los aspectos metodológicos del trabajo, como se abordó el problema, que modelo predictivo se utilizó y como se implementó. En este apartado se estará hablando de la propuesta del trabajo, como el desarrollo del trabajo se relaciona con el objetivo general del mismo y la evaluación de la solución presentada.

Implementación y mejora del modelo en la organización

El objetivo principal del trabajo fue predecir comportamientos transaccionales de la cartera asesorada de “Fintech Argentina S.A.” para poder detectar anomalías y definir alarmas. Con el Webservice “TP-MANA-Prediction-F” se pudo obtener la predicción para el 15/9 con un error cuadrático medio de 5 (menor a la media que se había obtenido para la base “tp_modelo”) para los primeros diez clientes de la misma. A medida que se vayan sumando más registros a la base predictiva se puede seguir entrenando el modelo para que el error medio siga disminuyendo.

Si bien el entrenamiento y aplicación del modelo se hicieron con solamente diez clientes y cuatro meses de operación, sirve de base para poder extenderlo al resto de la cartera asesorada. A su vez, en este caso se utilizó una base estática para entrenar al modelo obtenida de la devolución de una API para un momento en particular, por lo que a medida que pase el tiempo se puede ir alimentando la base de entrenamiento del modelo con los registros de los clientes que irán retornando la API. En base a este trabajo se puede plantear un proceso automático de recolección y actualización de datos para nutrir al modelo de la mejora manera.

A medida que se van detectando anomalías transaccionales en los clientes se pueden registrar e ingresar como input a la base para el re-entrenamiento de la misma, teniendo no solamente la cantidad de transacciones, sino sus variaciones y tipo de errores que tuvieron en el pasado. Realizando estas correcciones se podrían evitar los falsos positivos a futuro.

La aplicación del modelo puede mejorar la respuesta ante anomalías transaccionales disminuyendo el tiempo de respuesta ante contingencias y mejorando la calidad del servicio.

Oportunidades de mejora

Con las predicciones que se pudieron obtener con la implementación del modelo se puede establecer un sistema de alarmas que vaya comparando la cantidad de transacciones que va teniendo cada cliente contra la predicción con una varianza de 5 transacciones, y si lo que hizo el cliente en ese momento escapa a la franja predictiva se puede enviar una alerta al equipo de soporte técnico y comercial para que investiguen y accionen sobre la misma. El mismo script que se utilizó para obtener la predicción del modelo se puede utilizar como base para definir el sistema que vaya comparando las transacciones por hora.

Estas alertas pueden definirse no solamente para cantidad de transacciones, sino que también se puede utilizar la alerta para determinar la cantidad de dinero que representan dichas transacciones en el tiempo. También se pueden ir estableciendo alertas a medida que vayan surgiendo distintas necesidades de negocio. Particularmente para este caso se busca saber el cambio brusco en la cantidad de transacciones de un cliente, pero también se podría utilizar la misma información de la predicción para enviar alertas cuando un cliente supere determinada cantidad de transacciones o tenga un rendimiento sostenido en un período de tiempo en particular.

Al tener el sistema definido, tanto el equipo de negocio como de sistemas tiene la libertad de definir alertas que tengan sentido a la hora de buscar comportamientos o particularidades que se puedan atacar para mejorar la calidad del servicio.

Propuestas futuras

Como se habló en un principio, “Fintech Argentina S.A.” no solamente tiene operaciones en Argentina, sino que es una empresa en expansión en toda América Latina, por lo que la implementación de un modelo predictivo que ayude a mejorar la calidad del servicio en el resto de los países es una gran oportunidad. Con las predicciones de los clientes determinada, se puede desarrollar técnicamente el sistema que vaya comparando en tiempo real las predicciones con las transacciones que van ingresando para detectar y alertas anomalías en las mismas.

El modelo en sí es extrapolable a Chile y Uruguay por ejemplo, independientemente de las industrias que se manejen en dichos países, solamente se necesita nutrir al proyecto con la información de la cartera asesorada de los clientes correspondientes durante un tiempo de adaptación del mismo para que pueda funcionar de la mejor manera. A su vez, como trabajo futuro se puede analizar los componentes principales que hacen que una compañía presente

anomalías en sus transacciones, teniendo en cuenta cuestiones tanto particulares de la misma como las del entorno (situación país, estacionalidad, etc.).

En sí, la solución del trabajo se presenta como una integral en todo sentido al objetivo principal de “Fintech Argentina S.A.”. Con la implementación del modelo no solamente se puede predecir el comportamiento transaccional de los principales clientes de la cartera asesorada, sino que sirve como plataforma para poder buscar oportunidades de mejora para toda su cartera de clientes independientemente del país y la industria, así como también se puede utilizar la misma lógica para mejoras de procesos y recursos internos.

Conclusión

Con el desarrollo de este trabajo se buscó mejorar la calidad de servicio que “Fintech Argentina S.A.” le otorga a sus clientes. Para brindar una atención de calidad superadora a los clientes de la cartera asesorada, se debe responder de manera inmediata ante cualquier tipo de inconveniente que pueda tener el procesamiento de los pagos de un cliente en particular. Para lograr un accionar lo más rápido posible frente a estos problemas, ¿como se puede, analizando el comportamiento transaccional de la cartera asesorada, predecir el mismo y detectar anomalías?

Con el fin de poder responder esta incógnita se tomó el registro de las últimas transacciones que tuvieron los 10 principales clientes de dicha cartera en los últimos 4 meses, y se utilizó un modelo de machine learning para predecir que cantidad de transacciones deberían recibir los mismos en las horas/día siguiente. Se utilizó la predicción como base de un disparador de alertas en los casos que difiera mucho el incoming de las transacciones con respecto a la predicción obtenida.

El objetivo principal del trabajo se desglosó en base a este accionar, se obtuvo el histórico de los clientes (el cual tuvo un tratamiento previo a su análisis) y se utilizó el mismo para entrenar un modelo predictivo que entregue las bases para las futuras alertas, particularmente utilizando “Boosted Gradient Trees” como modelo predictivo principal sobre la herramienta de Azure Machine Learning obteniendo un error cuadrático medio de la predicción de 5 transacciones por hora. Se pudo concluir en que utilizando el modelo definido, se pueden disparar alertas de comportamiento anómalo cuando en una hora, un cliente procese más o menos 5 transacciones de las esperadas.

En los tres apartados se desarrolló con mayor detalle los aportes realizados en el trabajo. Se comenzó con la explicación de cómo la empresa “Fintech Argentina S.A.” abordaba esta problemática en el pasado, cual era el contexto en el que se encontraba y porque necesita un cambio de enfoque con respecto a su forma de trabajo y tratamiento de datos. Luego se detallaron los aspectos metodológicos del trabajo, como se consiguieron los datos, que tipo de modelo predictivo se utilizó, como se constituyó el script con el cual se trataron los datos y se comunicó con el Webservice definido para su predicción. Y finalmente se abordó la propuesta para la implementación del trabajo final y los beneficios que este conlleva.

Los aprendizajes de este trabajo se pueden extrapolar a cualquier empresa que provea servicios de procesamiento de pagos, independientemente del país en el que opere. La metodología y los pasos seguidos pueden utilizarse más allá de la industria, tomándose como “framework” de trabajo de análisis predictivo ante cualquier modelo de negocio. Por ejemplo: Predicción de cola de establecimiento médico en base a su histórico de reservas.

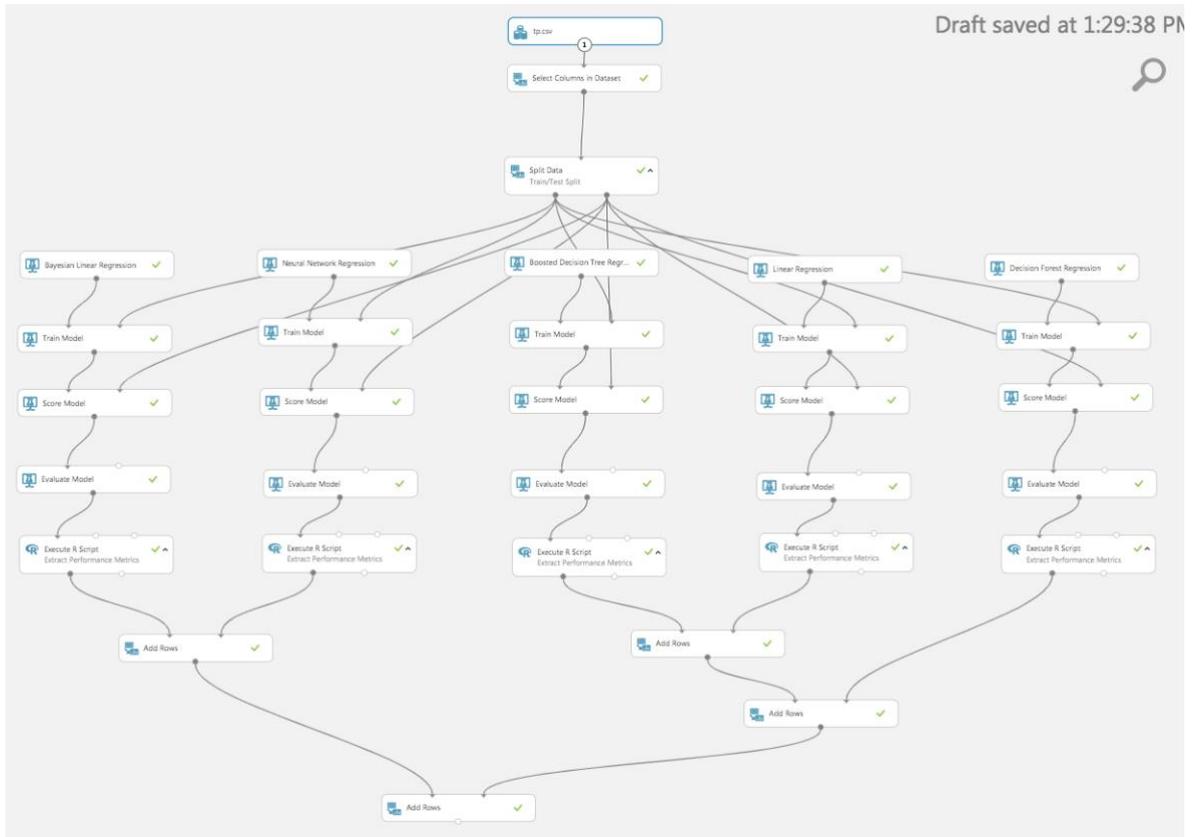
En base al desarrollo del trabajo se puede definir una línea de trabajo futura para definir un Webservice que vaya actualizando automáticamente su base de predicción con el tiempo, tomando como input el resultado de dicha predicción contra la realidad, y permita tener la información/capacidad suficiente para obtener predicciones con mayor anticipación.

Referencias bibliográficas

- Asli Demirgüç-Kunt, L. K. (2017). *Medición de la inclusión financiera y la revolución de la tecnología financiera*. World Bank Group.
- Chishti, S. y. (2016). *The fintech book. The financial technology handbook for investors, entrepreneurs and visionaries*. Great Britain: Wiley.
- CLIFTON PHUA1*, V. L. (2010). *A Comprehensive Survey of Data Mining-based Fraud Detection Research*.
- Company, M. a. (2019). *Global Payments Report 2019: Amid sustained growth, accelerating challenges demand bold actions*.
- Desai, F. (2015). *The Fintech boom and bank innovation*. Forbes.
- Devis, M. y. (2018). *Catching up with Indonesia's fintech industry*. Australia: Australian Centre For Financial Studies, University Of Melbourne.
- Fawcett, F. P. (2013). *DATA SCIENCE AND ITS RELATIONSHIP TO BIG DATA AND DATA-DRIVEN DECISION MAKING*. Mary Ann Liebert, Inc.
- Igual, D. (2016). *Fintech. Lo que la tecnología hace por las finanzas*.
- LaValle, S. (2012). *Big Data, Analytics and the Path from insights to value*. MITSloan.
- Lee, I. y. (2018). *Fintech: Ecosystem, business models, investment decisions, and challenges*. USA: Kelley School of Business, Indiana University.
- Monkol Lek, B. A. (2001). *DATA MINING PROTOTYPE FOR DETECTING E-COMMERCE FRAUD*.
- P. Berkmen, K. B. (2019). *Fintech in Latin America and the Caribbean: Stocktaking*. IMF Working Papers.
- Velarde, N. M. (2019). *Principales factores del desarrollo de la industria fintech para un país emergente en Latinoamérica*.
- Vives, X. (2017). *The Impact of Fintech on Banking*.

Anexos

- Comparación de modelos: Experimento de Azure machine learning que deja comparar modelos predictivos con la misma base. Nombre: Compare Regression Models.



Apéndices

- Respuesta de API: Es la información que devuelve internamente “Fintech Argentina S.A.” con respecto a la cantidad de transacciones que tiene un cliente en un período de tiempo. Se define en un csv.
 - Estructura del archivo:
 - ID: Identificador del cliente
 - Fecha: Fecha consultada. Formato: 9/15/19 0:00
 - Industria: Identificador de la industria a la cual pertenece el cliente.
- Valores:
- Apparel
 - Electro y electrónicos

- Home
 - Retail
 - Venta Directa
 - Ticket promedio: Valor promedio de las transacciones que tiene un cliente en general. Valores:
 - 0-1000
 - 1000-2500
 - 2500-5000
 - 5000-10000
 - 10000-15000
 - >15000
 - Q: Cantidad de transacciones
- Eventos: Base que indica si el día de la fecha es un evento importante para el e-commerce y el cálculo de cuantos días faltan para el próximo, así como también si es feriado o no. Estructura del archivo:
 - Fecha:
 - Feriado: Indica si es feriado. Valores: 1 y 0.
 - Evento: Indica si es un evento o no. Valores: 1 y 0.
 - Dias para el próximo evento: Indica cuantos días faltan para el próximo evento. Valores: 0-365.

Fecha	Feriado	Evento?	Dias para el proximo evento
4/1/19 0:00	0	0	42
4/2/19 0:00	1	0	41
4/3/19 0:00	0	0	40
4/4/19 0:00	0	0	39
4/5/19 0:00	0	0	38
4/6/19 0:00	0	0	37
4/7/19 0:00	0	0	36
4/8/19 0:00	0	0	35
4/9/19 0:00	0	0	34

- Tp_modelo_entrenamiento: Base que se utilizó para entrenar el modelo predictivo.
 - ID: Identificador del cliente
 - Day: Día de la fecha que se obtuvieron las métricas

- Month: Mes de la fecha que se obtuvieron las métricas
- Year: Año de la fecha que se obtuvieron las métricas
- Hora: Hora de la fecha que se obtuvieron las métricas
- Weekday: Día de la semana. Valores: 1-Lunes, 2-Martes, 3-Miercoles, 4-Jueves, 5-Viernes, 6-Sabado y 7-Domingo.
- a. Weekend?: Si es fin de semana. Valores: 0-No y 1-Si.
- b. “Es feriado?”. Valores: 0-No y 1-Si.
- c. “Está en algún evento?”. Valores: 0-No y 1-Si.
- d. “Cuantos días faltan para el próximo evento?”. Valores 0-365.
- e. Industria: Identificador de la industria a la cual pertenece el cliente.
Valores:
 1. Apparel
 2. Electro y electrónicos
 3. Home
 4. Retail
 5. Venta Directa
- f. Ticket promedio: Valor promedio de las transacciones que tiene un cliente en general. Valores:
 1. 0-1000
 2. 1000-2500
 3. 2500-5000
 4. 5000-10000
 5. 10000-15000
 6. >15000

g. Q: Cantidad de transacciones en el período

ID	Day	Month	Year	Hour	Weekday	Weekend?	Days to next	Event?	Industria	Ticket promé	Q	Holiday?
2	1	5	2019	0	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	1	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	2	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	3	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	4	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	5	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	6	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	7	3	0	12	0	Home	2500-5000	0	1
2	1	5	2019	8	3	0	12	0	Home	2500-5000	0	1

- Base_a_predecir: Base sin tratar para predecir. Estructura del archivo:

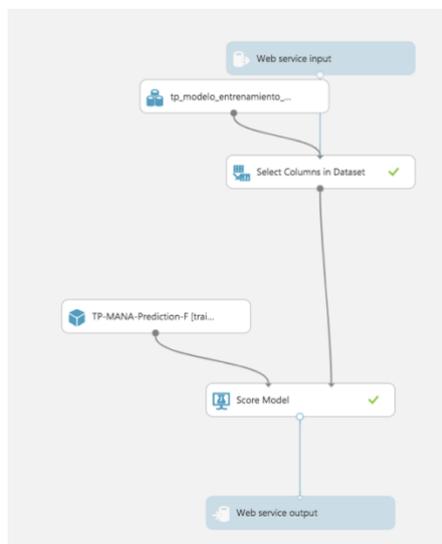
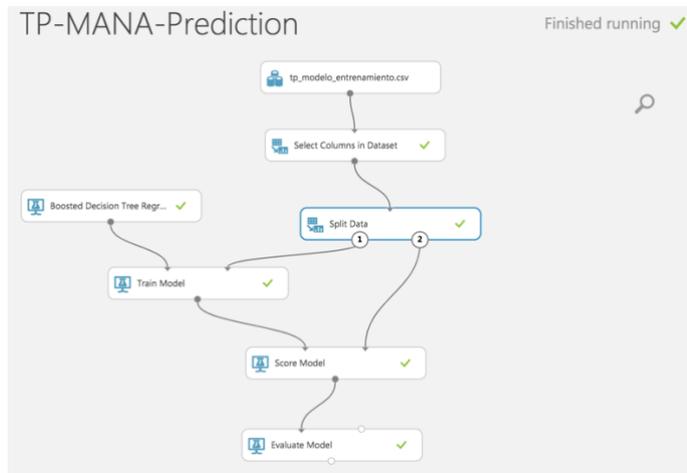
- ID: Identificador del cliente
- Fecha: Fecha consultada. Formato: 9/15/19 0:00
- Industria: Identificador de la industria a la cual pertenece el cliente.

Valores:

- Apparel
 - Electro y electrónicos
 - Home
 - Retail
 - Venta Directa
- Ticket promedio: Valor promedio de las transacciones que tiene un cliente en general. Valores:
 - 0-1000
 - 1000-2500
 - 2500-5000
 - 5000-10000
 - 10000-15000
 - >15000
 - Q: Vacío para la predicción

ID	Date	Q	Industria	Ticket promedio
2	9/15/19 0:00		Home	2500-5000
2	9/15/19 1:00		Home	2500-5000
2	9/15/19 2:00		Home	2500-5000
2	9/15/19 3:00		Home	2500-5000

- TP-MANA-Prediction: Modelo predictivo creado en Azure Machine Learning Studio y exposición como Web Service.



- Script (prepare_post_create.py): Script que toma la respuesta de la API, la modifica, le suma información de “eventos” y la convierte en base_a_predecir_adaptada para ir contra el web service de TP-MANA-Prediction y obtener la base con la predicción.

Código:

```
# Recursos
import datetime
from datetime import datetime
from datetime import time
import openpyxl
from openpyxl import Workbook
import xlrd
from collections import OrderedDict
import urllib.request
import json
```

```

# Imprimo la hora para ver cuanto tarda
print(datetime.now())

# -----

# Previo a comenzar, me traigo las fechas de los eventos del año

# Open the workbook and select the first worksheet
wb = xlrd.open_workbook('eventos.xlsx')
sh = wb.sheet_by_index(0)
# List to hold dictionaries
events_list = []
# Iterate through each row in worksheet and fetch values into dict
for rownum in range(2, sh.nrows):
    event = OrderedDict()
    row_values = sh.row_values(rownum)
    event['fecha'] = row_values[0]
    event['holiday'] = row_values[1]
    event['event?'] = row_values[2]
    event['next_event'] = row_values[3]
    events_list.append(event)
# Serialize the list of dicts to JSON
j = json.dumps(events_list)
eventos_j=json.loads(j)

# -----

# 1-Abro el excel
wb=openpyxl.load_workbook("base_a_predecir.xlsx")

# Me paro sobre la hoja de q
sheet=wb.get_sheet_by_name("Q")

# Tomo el valor de una celda
# valor=sheet['A1'].value

# Tomo el valo de una celda
# sheet.cell(row=1,column=3).value

# Numero de filas
filas=sheet.max_row+1

# 2-Agrego las columnas que faltan
sheet['F1'].value="Year"
sheet['G1'].value="Month"

```

```

sheet['H1'].value="Day"
sheet['I1'].value="Hour"
sheet['J1'].value="Weekday"
sheet['K1'].value="Weekend?"
sheet['L1'].value="Holiday?"
sheet['M1'].value="Event?"
sheet['N1'].value="Next_event"

```

```
# 3-Loopeo todas las filas
```

```
for x in range(2,filas):
```

```
    date_obj=sheet.cell(row=x,column=2).value
```

```
    # 3.1-Tomo año, mes, dia y hora de la fecha que viene del excel
```

```
    year=date_obj.date().year
```

```
    month=date_obj.date().month
```

```
    day=date_obj.date().day
```

```
    hour=date_obj.time().hour
```

```
    # 3.2-Digo que dia de la semana es
```

```
    weekday=date_obj.isoweekday()
```

```
    # 3.3-Digo si es fin de semana
```

```
    if weekday==6 or weekday==7:
```

```
        weekend=1
```

```
    else:
```

```
        weekend=0
```

```
        pass
```

```
    # 3.4-Me fijo si hay algun evento y me traigo cuanto falta para el próximo consultando al json de "j_eventos"
```

```
    for y in eventos_j:
```

```
        date_evento=datetime.fromtimestamp((y['fecha']-25568)*86400.0)
```

```
        if date_evento.date()==date_obj.date():
```

```
            # 4-Pongo los valores en las columnas correspondientes
```

```
            sheet.cell(row=x,column=6).value=year
```

```
            sheet.cell(row=x,column=7).value=month
```

```
            sheet.cell(row=x,column=8).value=day
```

```
            sheet.cell(row=x,column=9).value=hour
```

```
            sheet.cell(row=x,column=10).value=weekday
```

```
            sheet.cell(row=x,column=11).value=weekend
```

```

        sheet.cell(row=x,column=12).value=y['holiday']
        sheet.cell(row=x,column=13).value=y['event?']
        sheet.cell(row=x,column=14).value=y['next_event']

    pass

pass

pass

# 6- Guardo y creo un nuevo excel para dejarlo como log
wb.save("base_a_predecir_adaptada.csv")

print("Se creó el .csv que se va a utilizar para la predicción")

# 7-Una vez tengo el excel con toda la info lo transformo en un json par enviarlo por POST
al Webservice
# Open the workbook and select the first worksheet
wb = xlrd.open_workbook('base_a_predecir_adaptada.csv')
sh_final = wb.sheet_by_index(0)

# List to hold dictionaries
payments_list = []
# Iterate through each row in worksheet and fetch values into dict -> Armo el array y
acomodo las posiciones según lo que necesita el Webservice
# Mejora a futuro-> Que coincidan las posiciones
for rownum in range(2, sh_final.nrows):
    payment = OrderedDict()
    row_values = sh_final.row_values(rownum)
    payment['ID'] = int(row_values[0])
    payment['Day'] = int(row_values[7])
    payment['Month'] = int(row_values[6])
    payment['Year'] = int(row_values[5])
    payment['Hour'] = int(row_values[8])
    payment['Weekday'] = int(row_values[9])
    payment['Weekend?'] = int(row_values[10])
    payment['Days to next event'] = int(row_values[13])
    payment['Event?'] = int(row_values[12])
    payment['Industria'] = row_values[3]
    payment['Ticket promedio'] = row_values[4]
    payment['Q'] = row_values[2]
    payment['Holiday?'] = int(row_values[11])

    payments_list.append(payment)

# Serialize the list of dicts to JSON
j2 = json.dumps(payments_list)

```

```

payments_j=json.loads(j2)

print("Se creó el json con la base a predecir")

# -----

# 8-Una vez que tengo el json de la base hago el POST contra el Webservice

data = {
    "Inputs": {
        "input1": payments_j
    },
    "GlobalParameters": {
    }
}

body = str.encode(json.dumps(data))

url =
'https://ussouthcentral.services.azureml.net/workspaces/fe4f5791e4ad4bca98491179f3d0b2
04/services/740b04f43a874f5db243b6d259d6c778/execute?api-
version=2.0&format=swagger'
api_key =
'uCJ94MfnlZrRsrcHdgxdpg7BCpsk370uExMLqgJMikIigBojzKFWyzboowRojc3aEVRbx
/6I0uBVTlq3SswsXg== ' # Replace this with the API key for the web service
headers = {'Content-Type':'application/json', 'Authorization':('Bearer '+ api_key)}

req = urllib.request.Request(url, body, headers)

try:
    response = urllib.request.urlopen(req)

    result = response.read()
    print("Se ejecutó con éxito el Webservice")
except urllib.error.HTTPError as error:
    print("The request failed with status code: " + str(error.code))

    # Print the headers - they include the request ID and the timestamp, which are useful for
    debugging the failure
    print(error.info())
    print(json.loads(error.read().decode("utf8", 'ignore'))))

# -----
-----

```

9- Una vez obtengo la respuesta del Webservice me imprimo la predicción en un .json y en un .xlsx

```
# 9.1-Genero el .json
hora=str(datetime.now())
prediction = open("prediction_"+hora+".json", "w")
```

```
# Ordeno el json antes de imprimirlo
my_json = result
parsed = json.loads(my_json)
```

```
# Imprimo la info en el .json
prediction.write(json.dumps(parsed, indent=4, sort_keys=True))
prediction.close()
```

```
print("se guardó el .json con la predicción")
```

```
# -----
```

9.2 Genero el excel y guardo la info

9.2.1- Creo el excel

```
wb_prediction=Workbook()
# select prediction.xlsx
sheet_prediction=wb_prediction.active
```

9.2.2- Creo las columnas

```
sheet_prediction.cell(row=1,column=1).value="ID"
sheet_prediction.cell(row=1,column=2).value="Day"
sheet_prediction.cell(row=1,column=3).value="Month"
sheet_prediction.cell(row=1,column=4).value="Year"
sheet_prediction.cell(row=1,column=5).value="Hour"
sheet_prediction.cell(row=1,column=6).value="Weekday"
sheet_prediction.cell(row=1,column=7).value="Weekend?"
sheet_prediction.cell(row=1,column=8).value="Days to next event"
sheet_prediction.cell(row=1,column=9).value="Event?"
sheet_prediction.cell(row=1,column=10).value="Industria"
sheet_prediction.cell(row=1,column=11).value="Ticket Promedio"
sheet_prediction.cell(row=1,column=12).value="Holiday?"
sheet_prediction.cell(row=1,column=13).value="Q/Scored Label"
```

9.2.3- Recorro el json y lo imprimo en el excel

```
result=json.loads(result)
result_prediction=result['Results']['output1']
```

```
x=2
```

```
for result_p in result_prediction:
```

```

# Tomo los valores del json
cust_p=result_p['ID']
day_p=result_p['Day']
month_p=result_p['Month']
year_p=result_p['Year']
hour_p=result_p['Hour']
weekday_p=result_p['Weekday']
weekend_p=result_p['Weekend?']
days_p=result_p['Days to next event']
event_p=result_p['Event?']
industria_p=result_p['Industria']
ticket_p=result_p['Ticket promedio']
holiday_p=result_p['Holiday?']
tpn_p=result_p['Scored Labels']

# Los imprimo en el excel
sheet_prediction.cell(row=x,column=1).value=cust_p
sheet_prediction.cell(row=x,column=2).value=day_p
sheet_prediction.cell(row=x,column=3).value=month_p
sheet_prediction.cell(row=x,column=4).value=year_p
sheet_prediction.cell(row=x,column=5).value=hour_p
sheet_prediction.cell(row=x,column=6).value=weekday_p
sheet_prediction.cell(row=x,column=7).value=weekend_p
sheet_prediction.cell(row=x,column=8).value=days_p
sheet_prediction.cell(row=x,column=9).value=event_p
sheet_prediction.cell(row=x,column=10).value=industria_p
sheet_prediction.cell(row=x,column=11).value=ticket_p
sheet_prediction.cell(row=x,column=12).value=holiday_p
sheet_prediction.cell(row=x,column=13).value=tpn_p

x=x+1
pass

wb_prediction.save("prediction_"+hora+".xlsx")
print("Se guardó el excel con la predicción")

print(datetime.now())

```