

Propuesta de Trabajo Profesional

Análisis de los factores que influyen en la
retención de usuarios para la aplicación móvil y web de
Restorando

Autores

Apellido, Nombre	E-Mail	Padrón
Collinet, Jorge	collinetjorge@gmail.com	92131
Dausá, Juan Ángel	juandausa@gmail.com	92267
Sella Faena, Jasmina	jasminasf@gmail.com	91958

Carrera

Ingeniería en Informática

Director

Lic. Luis Argerich

Objetivo del Proyecto

Visión del Proyecto

Análisis de Mercado y Motivación

Alcance

Requerimientos

Productos a Generar

Herramientas de Desarrollo

Hardware

Software

Metodología

Estimación de Esfuerzo

Esfuerzo de Desarrollo

Esfuerzo del proyecto

Plan de Trabajo

Currículum de los Autores

Collinet, Jorge Francisco

Dausá, Juan Ángel

Sella Faena, Jasmina

Plan de Cursado

Collinet, Jorge Francisco (Orientación: Gestión Industrial de Sistemas)

Dausá, Juan Ángel (Orientación: Gestión Industrial de Sistemas)

Sella Faena, Jasmina (Orientación: Gestión Industrial de Sistemas)

Objetivo del Proyecto

El proyecto tiene como objetivo predecir, en las primeras semanas de interacción de un usuario con la plataforma de Restorando, si el primero será un usuario activo o no. Cada predicción deberá explicar cual es el comportamiento del usuario que se correlaciona con la predicción hecha.

Para esto es necesario poner en claro cuál es la definición de un usuario activo para Restorando. Se considera un usuario activo, uno que haya realizado al menos una reserva en los últimos 30 días corridos. Su opuesto es considerado usuario inactivo, no retenido o *churn*.

Una definición más amplia de esto puede verse en la siguiente clasificación:

- **Usuario Nuevo:** Usuarios que no realizaron su primer reserva dentro de sus primeros 30 días.
- **Usuario Activo:** Usuario que posee al menos una reserva en los últimos 30 días. A su vez, pueden dividirse en:
 - **Usuarios activados:** Usuarios que hicieron su primer reservan en los últimos 30 días.
 - **Usuarios Retenidos:** Usuarios que hicieron al menos una reserva en los últimos 30 días y no es su primera.
- **Churn:** Usuarios que no hicieron ninguna reserva en los últimos 30 días.

Visión del Proyecto

Con el desarrollo de este proyecto se pretende lograr a futuro una reducción de la cantidad de usuarios que dejan de usar la plataforma de Restorando. Hoy en día, los usuarios que abandonan la plataforma constituyen el 98% de los usuarios totales. Se busca contribuir al análisis de las causas de este suceso y colaborar en la búsqueda de una solución para evitarlo.

Análisis de Mercado y Motivación

En la actualidad, las métricas relacionadas a las ganancias están fuertemente ligadas a la retención de usuarios, es decir, a la cantidad de usuarios que vuelven a usar una aplicación o sitio web luego de un periodo de tiempo determinado. Su opuesto es llamado *churn* y consiste en la cantidad de usuarios que abandonan la plataforma en un periodo de tiempo dado.

El problema de negocio al que se enfrentan las compañías es cómo aumentar la retención de usuarios y que no haya *churn*. Por ello, usualmente dan incentivos y tratos especiales a los usuarios que tienen más probabilidades de convertirse en *churn*, para evitar que esto ocurra. Estos incentivos tienen un costo por lo cual es necesario diferenciar a aquellos usuarios a los que es necesario aplicar un trato especial de los que no.

Si Restorando pudiera predecir qué usuarios nuevos van a ser activos y cuales no, o cuales de los usuarios activos se retienen y cuales no, podría entender mejor qué tipos de usuario posee, qué acciones son clave en cada caso y definir tratos especiales para cada usuario y aplicarlos en tiempo real.

La principal motivación que encontramos, para realizar el proyecto propuesto, es poder trabajar con grandes volúmenes de datos obtenidos como resultado de la interacción de una empresa real con sus usuarios. A su vez, analizar la información de los mismos con el objetivo de obtener perfiles de comportamiento.

Con esto en mente, será posible proponer mejoras en los sistemas actuales, favoreciendo una mejor interacción con los usuarios e intentando mantener la mayor cantidad de perfiles activos.

Alcance

Requerimientos

El trabajo consiste en un análisis incremental de los datos generados en las distintas aplicaciones web y mobile de Restorando, intentando averiguar y predecir el comportamiento de los usuarios. Dado que no se conocen en detalle todos los datos ni todas las interacciones posibles comenzaremos haciendo un análisis manual de la información.

El requerimiento inicial es poder predecir en las primeras dos semanas de interacción del usuario con el sistema si el mismo será un usuario activo o no. Para ello, procederemos a hacer un análisis del comportamiento de los usuarios que realizan reservas de restaurantes en la plataforma y se mantienen activos, en contraste con aquellos usuarios que dejan de usarla.

Productos a Generar

- Dashboards que permitan visualizar los datos recolectados por Restorando en sus logs.
- Simulaciones que permitan comprender el comportamiento de los usuarios.
- Modelos de Machine Learning que sirvan para predecir la probabilidad de churn en base a los datos que se tienen del usuario.
- Notebook en Jupyter¹ con los algoritmos principales y sus resultados.

Herramientas de Desarrollo

Hardware

Las especificaciones mínimas del equipo de desarrollo son:

Parte	Detalle
Procesador	X64 Doble Nucleo
RAM	4GB
HD Space	16GB
OS	Windows 7/Linux

Software

El lenguaje de desarrollo inicial será Python² en su versión 2.7, se trata un lenguaje de tipado dinámico e interpretado que es fácil de usar y mantener, lo que ayuda a disminuir el esfuerzo de desarrollo en gran medida. Sus librerías son de código abierto y tienen soporte constante de la comunidad open source.

Sin embargo, existe la posibilidad de usar otro lenguaje dependiendo de la necesidad que se presente para el análisis de datos.

La presentación de los datos obtenidos se realizará, cuando sea posible, utilizando Bokeh³, una librería de visualización interactiva de Python que permite ver los gráficos obtenidos en formato .html en el navegador.

¹ Jupyter: <http://jupyter.org/>

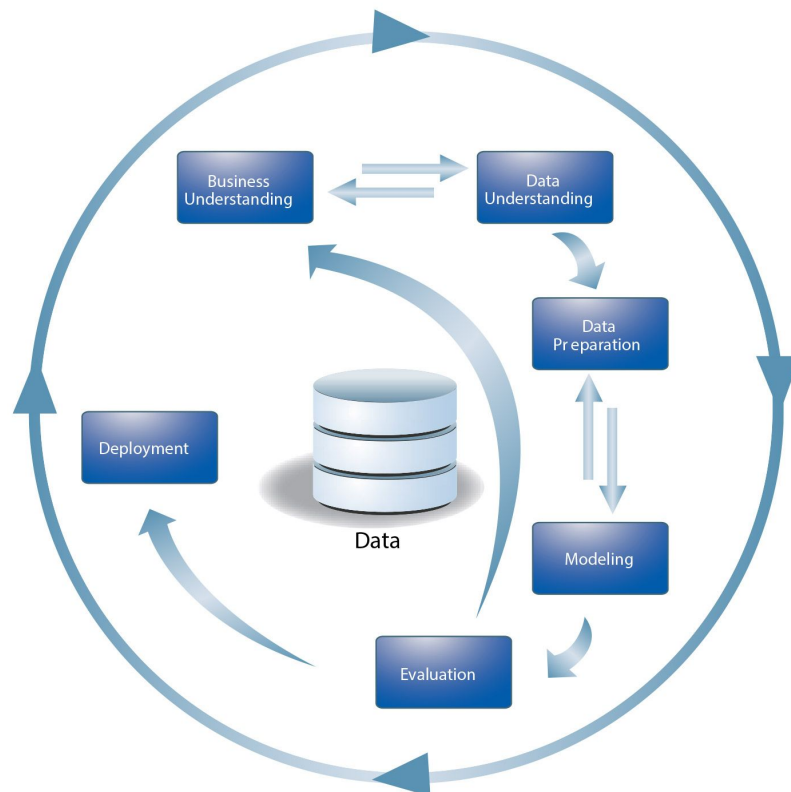
² Python: <https://www.python.org/>

³ Bokeh: <http://bokeh.pydata.org/en/latest/>

En caso de ser necesario se utilizará el motor de bases de datos MySql⁴. Se ha optado por este por ser gratuito, multiplataforma y de público conocimiento en el ambiente IT.

Metodología

Se realizarán ciclos iterativos en base a la metodología CRISP⁵, la cual se utiliza en proyectos de implementación de minería de datos y está estructurada en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.



Esta metodología será acompañada por un seguimiento quincenal con análisis de resultados y control de objetivos. Esto permitirá darle a conocer avances al cliente, logrando una comunicación fluida con el mismo y su participación. Como consecuencia, se obtiene una rápida corrección ante desvíos y un preciso entendimiento del problema y sus variables.

⁴ MySql: <https://www.mysql.com/>

⁵ CRISP: https://es.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Estimación de Esfuerzo

Dado que es una propuesta abierta se plantea el desarrollo en iteraciones. El contenido de las mismas podrá variar de acuerdo a lo realizado en las iteraciones anteriores. A pesar de esto se plantea un esquema general para poder encuadrar el desarrollo en el formato del Trabajo Profesional.

Siguiendo el esquema propuesto por CRISP se plantean dos iteraciones, más cortas al resto, para lograr un buen entendimiento del modelo de datos y de la metodología.

Esfuerzo de Desarrollo

ID	Descripción	Esfuerzo (hs)
1	Procesamiento de datos	32
1.1	Modulo de lectura de datos	10
1.2	Modulo de normalización de datos	12
1.3	Modulo de graficación de resultados	10
2	Iteración 1	90
2.1	Comprensión del negocio y de los datos	
2.1.1	Entendimiento del negocio	30
2.1.2	Exploración inicial de los datos	40
2.1.3	Análisis de completitud de los datos	10
2.1.4	Análisis de calidad de los datos	10
3	Iteración 2	100
3.1	Métricas básicas	
3.1.1	Definición de métricas básicas a implementar	10

3.1.2	Priorización de las métricas	6
3.1.3	Implementación de métricas	50
3.2	Testing	
3.2.1	Definición de casos de prueba	12
3.2.2	Ejecución de casos de prueba	6
3.3	Presentación de datos	
3.3.1	Graficación de resultados	6
3.3.2	Análisis de Resultados	10
4	Iteración 3	180
4.1	Métricas	
4.1.1	Definición de métricas a implementar	20
4.1.2	Priorización de las métricas	6
4.1.3	Implementación de métricas	90
4.2	Testing	
4.2.1	Definición de casos de prueba	18
4.2.2	Ejecución de casos de prueba	10
4.3	Presentación de datos	
4.3.1	Graficación de resultados	16
4.3.2	Análisis de Resultados	20
5	Iteración 4	180
5.1	Métricas	
5.1.1	Definición de métricas a implementar	20
5.1.2	Priorización de las métricas	6

5.1.3	Implementación de métricas	90
5.2	Testing	
5.2.1	Definición de casos de prueba	18
5.2.2	Ejecución de casos de prueba	10
5.3	Presentación de datos	
5.3.1	Graficación de resultados	16
5.3.2	Análisis de Resultados	20
6	Iteración 5	200
6.1	Comprensión del negocio y de los datos	
6.1.1	Exploración de los datos de las plataformas mobile	20
6.2	Métricas	
6.2.1	Definición de métricas a implementar	20
6.2.2	Priorización de las métricas	6
6.2.3	Implementación de métricas	90
6.3	Testing	
6.3.1	Definición de casos de prueba	18
6.3.2	Ejecución de casos de prueba	10
6.4	Presentación de datos	
6.4.1	Graficación de resultados	16
6.4.2	Análisis de Resultados	20
7	Iteración 6	180
7.1	Métricas	
7.1.1	Definición de métricas a implementar	20

7.1.2	Priorización de las métricas	6
7.1.3	Implementación de métricas	90
7.2	Testing	
7.2.1	Definición de casos de prueba	18
7.2.2	Ejecución de casos de prueba	10
7.3	Presentación de datos	
7.3.1	Graficación de resultados	16
7.3.2	Análisis de Resultados	20
8	Iteración 7	90
8.1	Discusión y presentación de los resultados	
8.1.1	Análisis de los resultados finales obtenidos	40
8.1.2	Gráficos globales de los resultados	20
8.1.3	Confección de Informe final para presentar al cliente	30
	TOTAL	1052

Esfuerzo del proyecto

ID	Descripción	Esfuerzo (hh)
1	Análisis de requerimientos	74
1.1	Identificación de requerimientos funcionales	10
1.2	Identificación de requerimientos no funcionales	10
1.3	Análisis de requerimientos	40
1.4	Estimación requerimientos	14
2	Configuración de ambientes	21
2.1	Configuración de repositorio de versionado	1
2.2	Instalación y configuración de entorno de desarrollo	8
2.3	Instalación y configuración de herramienta de generación de gráficos	12
3	Desarrollo	1052
4	Documentación	46
4.1	Generación de documentación técnica	32
4.2	Redacción de un manual de usuario	14
5	Gestión	175
5.1	Gestión de requerimientos, alcance y cambios	20
5.2	Gestión de riesgos	20
5.3	Planificación y calendarización	50
5.4	Seguimiento de avance	32
5.5	Gestión de entregas incrementales	32
5.6	Gestión de la calidad del producto	21
6	Presentación	40
	TOTAL	1408

Plan de Trabajo

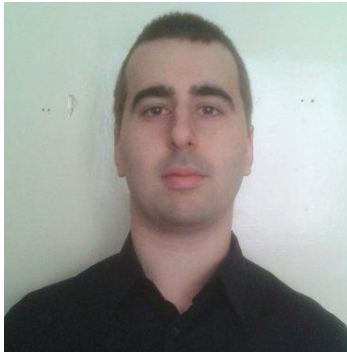
Se estima que el trabajo completo a realizar tomará 6 meses, con una dedicación de 78 hs mensuales al desarrollo del trabajo de parte de cada integrante. El proceso de desarrollo será iterativo e incremental, con dos iteraciones iniciales de dos semanas cada una y cinco iteraciones de un mes, a lo largo de las cuales se efectuarán tanto tareas de análisis, desarrollo, testing y documentación como tareas de gestión, cada una con distinta carga horaria dependiendo de la etapa del proyecto.

Para la gestión del proyecto, se adoptarán algunas prácticas de metodologías ágiles. Como en Scrum, se realizarán reuniones de Planificación, Revisión y Retrospectiva para cada iteración y una reunión informal diaria de reporte de estado que se llevará a cabo tanto presencialmente como de forma remota dependiendo de la disponibilidad de los involucrados. Se utilizará un tablero de Kanban para organizar el trabajo en cada iteración.

Se efectuará la gestión de riesgos con una planilla en donde se registrarán los riesgos identificados, junto con su probabilidad de manifestación, impacto sobre el objetivo del proyecto y su exposición estimada. Se describirán también los planes de respuesta y planes de contingencia para cada riesgo.

Se buscará que el contacto con el tutor y el cliente se efectúen principalmente durante las reuniones de Revisión, aunque se intentará integrar al cliente al proceso de desarrollo día a día para facilitar la comunicación de requerimientos y evitar el retrabajo. Durante las reuniones de Revisión, se presentarán tanto al tutor como al cliente el avance alcanzado durante la iteración anterior. Se definirán prioridades, nuevos requerimientos y cambios en la planificación.

Currículum de los Autores



Collinet, Jorge Francisco

Información Personal **DNI:** 35.571.842
Fecha de Nacimiento: 10 de abril de 1991
Estado Civil: Soltero
Nacionalidad: Argentina
Teléfono: +54 9 11 4773 4138
E-mail: collinetjorge@gmail.com

Educación

Facultad de Ingeniería de la UBA, CABA

Ingeniería en Informática. *Esperado: 2016*

- Todas las materias cursadas.
- Promedio: 7

San Martin de Tours, Palermo, Ciudad Autónoma de Buenos Aires

Título Secundario, 2008

- Bachiller
- Promedio: 7

Idiomas

Inglés - Nivel Intermedio

Experiencia Laboral

Desarrollador de software en Baufest *Mayo 2015 - Febrero 2016*

- Desarrollo Java
- Gestión de bases de datos MySQL, SQL Server.
- Despliegue de productos en servidores como IIS, Apache, Tomcat.

Desarrollador de software en Navent *Febrero 2016 - actualidad*

- Desarrollo Java, Scala, Akka
- Gestión de bases de datos MySQL, Cassandra
- Despliegue de productos en servidores como IIS, Apache, Tomcat, Jetty

Conocimientos

Programación

- **Lenguajes:** C, C++, Python, JavaScript, Java, Scala
- **Frameworks:** Spring, Android SDK, Storm, Kafka
- **Base de Datos:** MySQL, Cassandra, Mongo db
- **Integración Continua:** Jenkins.



Dausá, Juan Ángel

Información Personal **DNI:** 35.347.161
Fecha de Nacimiento: 27 de marzo de 1990
Estado Civil: Soltero
Nacionalidad: Argentina
Teléfono: +54 9 11 6610 5135
E-mail: juandausa@gmail.com

Educación **Facultad de Ingeniería de la UBA**, Ciudad Autónoma de Buenos Aires
Ingeniería en Informática. *Esperado: 2016*

- Actualmente cursando el último cuatrimestre de la carrera.
- Promedio: 8

Instituto Industrial Luis A. Huego, San Telmo, Ciudad Autónoma de Buenos Aires
Título Secundario, 2009

- Técnico Electrónico con orientación en computadoras
- Promedio: 7

Idiomas **Inglés** - Nivel Intermedio

Experiencia Laboral **Full Stack Developer en Quadion** *Mayo 2014 - actualidad*

- Desarrollo backend .Net, Symfony y Java
- Desarrollo frontend web y mobile con HTML5, CSS3, JavaScript y JQuery, con la utilización de frameworks como Angular.
- Gestión de bases de datos MySQL, SQL Server.
- Despliegue de productos en servidores como IIS, Apache, Tomcat.

Conocimientos **Programación**

- **Lenguajes:** C, C++, C#, PHP, Python, JavaScript, Ruby, Java.
- **Frameworks:** .Net, Spring, Rails, Symfony, Android SDK.
- **Base de Datos:** MySQL, SQL Server, RocksDB.
- **Maquetado:** HTML5, CSS3, Sass..
- **Herramientas de Virtualización:** Docker, VirtualBox, VMWare.
- **Despliegue:** ISS, Apache, Tomcat.
- **Integración Continua:** Jenkins, Travis.
- **Sistema Operativo:** Linux, Windows.



Sella Faena, Jasmina

Información Personal **DNI:** 36.661.164
Fecha de Nacimiento: 7 de julio de 1990
Estado Civil: Soltera
Nacionalidad: Argentina
Teléfono: +54 9 11 3167 2301
E-mail: jasminasf@gmail.com

Educación **Facultad de Ingeniería de la UBA**, Ciudad Autónoma de Buenos Aires
Ingeniería en Informática. *Esperado: 2016*

- Actualmente cursando el último cuatrimestre de la carrera.
- Promedio: 7

Escuela Superior de Comercio Carlos Pellegrini, Recoleta, Ciudad Autónoma de Buenos Aires
Título Secundario, 2008

- Perito Mercantil
- Promedio: 7

Idiomas **Inglés** - Nivel Avanzado
Cambridge First Certificate in English

- Año 2007
- Grade B

Italiano - Nivel Básico

Experiencia Laboral **Developer en FDV Solutions** *Abril 2015 - actualidad*

- Desarrollo backend usando SpringMVC y Java.
- Desarrollo frontend web en php, HTML5 y JavaScript.
- Gestión de bases de datos MySQL, PostgreSQL.

Conocimientos **Programación**

- **Lenguajes:** Python, Java, C, C++, PHP, JavaScript, Android.
- **Frameworks:** Hibernate, JPA, Spring, Symfony, Android SDK.
- **Base de Datos:** MySQL, PostgreSQL.
- **Maquetado:** HTML5
- **Herramientas de Virtualización:** Docker, VirtualBox
- **Integración Continua:** Travis.
- **Sistema Operativo:** Linux.

Plan de Cursado

Collinet, Jorge Francisco (Orientación: Gestión Industrial de Sistemas)

1er Cuatrimestre 2016

Todas las materias de la carrera ya aprobadas. Durante el cuatrimestre se dedicará a la realización del trabajo profesional con el objetivo de recibirse.

Dausá, Juan Ángel (Orientación: Gestión Industrial de Sistemas)

1er Cuatrimestre 2016

[75.52] Taller de Programación II

Sella Faena, Jasmina (Orientación: Gestión Industrial de Sistemas)

1er Cuatrimestre 2016

[75.52] Taller de Programación II

[66.69] Criptografía y seguridad informática

[75.65] Manufactura Integrada por Computadora I

[75.70] Sistemas de programación no convencional de robots.