

# Artículo final

## Diseño e Innovación de productos en TI

Betgram  
Ingeniería de Sistemas y Computación,  
Universidad de los Andes

Noviembre 20, 2015

### Resumen

La práctica profesional del fútbol moviliza una cantidad significativa de atención y recursos en una parte importante del mundo globalizado. Alrededor de esta práctica se han construido industrias multimillonarias con requerimientos y stakeholders de la más diversa índole. En particular, la industria de la apuestas en línea se ha consolidado como una industria basada en la gestión del conocimiento y de reciente expansión en los mercados emergentes.

Como cabría esperar, la práctica profesional del fútbol genera grandes volúmenes de información en la web. Por una parte, la repetición de los encuentros deportivos y la participación constante de la mayoría de los competidores, deja tras su paso, un gran almacén de información estructurada como estadísticas de los encuentros, los equipos, los individuos, las condiciones del encuentro y las faltas o puntos de los participantes. Por otra parte, la difusión de contenidos en redes sociales ofrece una ventana a información no estructurada pero en cualquier caso actual, omnipresente y de alto potencial explicativo.

En el caso de una industria basada en la gestión del conocimiento, el análisis orientado a la predicción de este rico flujo de información ofrece grandes oportunidades para la generación de valor. El propósito de *betgram* es ofrecer a los apostadores futbolísticos una manera transparente de aprovechar de manera integral la información disponible en la web para realizar apuestas mejor informadas con mayor probabilidad de retorno.

**Palabras clave:** fútbol, predicción, apuestas, inteligencia social

## Motivación y experiencias previas

A pesar de la reciente explosión en la popularidad de las redes sociales y de su expansión a prácticamente todos los ámbitos de la actividad humana, su contenido permanece en gran medida sin ser explotado para el entendimiento de fenómenos sociales (Asur y Huberman 2010). Cabe preguntarse entonces, ¿Cómo puede ser utilizado el contenido derivado de redes sociales para la predicción de eventos de la vida real?

Las experiencias previas en este campo son limitadas pero presentan resultados alentadores. Por ejemplo, Asur Y Huberman (ibíd.) utilizaron métricas de popularidad en Twitter para predecir con éxito la recaudación en taquilla de películas de la industria cinematográfica próximas a ser estrenadas. En el ámbito deportivo, Sinha et al. (2011) exploran el poder predictivo de múltiples modelos para la Liga Nacional de Fútbol Americano de los Estados Unidos (NFL). Los resultados sugieren que modelos simples construidos a partir de la información disponible en Twitter tienen tanta capacidad predictiva como sus análogos más complejos construidos a partir de estadísticas históricas de desempeño deportivo. Adicionalmente, Yu y Kak (2012) presentan una revisión de los principales ejercicios predictivos que se han adelantado utilizando redes sociales. Los autores coinciden en señalar el amplio potencial de este tipo de ejercicios en campos tan distintos como la macroeconomía, el marketing, las industria cinematográfica y por supuesto el deporte.

Ahora bien, las experiencias previas identificadas se han desarrollado como parte de ejercicios académicos en el campo del aprendizaje de máquinas y la inteligencia artificial. En particular, no se tiene conocimiento de ninguna iniciativa de predicción utilizando inteligencia social que haya sido puesta a disposición de los apostadores deportivos. Este es el corazón de la propuesta de valor de *betgram* y es precisamente en este sentido que nuestro trabajo es único respecto a otras iniciativas de asesoramiento de apuestas deportivas disponibles en el mercado.

## Descripción

*betgram* es una aplicación web que ofrece a sus usuarios la posibilidad de realizar apuestas exitosas de manera transparente, conveniente y divertida. Betgram ofrece a sus usuarios la posibilidad de acceder a los mejores (a la luz de predicciones socialmente informadas) ofertas de los principales bookmakers de la web en tiempo aproximadamente real al costo mínimo de autenticarse y hacer algunos clics. Asimismo, Betgram hace uso de su rico stock de información para ofrecer todo tipo de estadísticas e infografías de alta calidad personalizables a la voluntad y las preferencias del usuario. A manera complementaria, Betgram sostiene un blog que sirve como punto de encuentro entre la aplicación y su comunidad online.

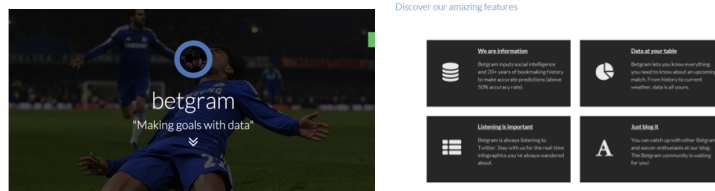


Figura 1: Página de inicio betgram (izquierda) y features relevantes (derecha)

*betgram* maneja un modelo de negocio freemium. Bajo la categoría free, los usuarios tienen la posibilidad de acceder al stock de información, estadísticas e infografías de Betgram y a algunos pocos picks relevantes para los partidos a disputarse próximamente. Bajo el modelo premium *betgram* ofrece a sus usuarios la posibilidad de personalizar su experiencia con la información, estadísticas e infografías y la posibilidad de acceder a los picks de todos los partidos a disputarse próximamente.

## Experiencia de utilización

El diseño del sitio web de *betgram* está orientado a atraer la atención de los usuarios de una manera agradable y sencilla. El principal atractivo de diseño es la disposición de la información, la cual busca documentar de manera clara datos relevantes para nuestros usuarios, logrando una aplicación web que en la que se encuentra información relevante y en tiempo real.

*betgram* logra diferenciarse de las otras alternativas al ser una aplicación web que brinda una experiencia informativa, gráfica y agradable a fans de fútbol, proporcionando contenido ilustrativo relevante. Como se mencionó anteriormente, la experiencia del usuario está acotada por su categoría (free o premium).

Uno de los factores diferenciadores más importantes de *betgram* es el la obtención y análisis de información social relacionada con la Premier League. Es por esto que el movimiento en las redes sociales se considera como fuente valiosa de recursos. En la siguiente imagen se muestra la fuerza que tiene cada uno de los clubes de la Premier League en twitter, todos los datos son obtenidos en tiempo real y actualizados automáticamente en la plataforma de *betgram*.

*betgram* busca acoplar todo tipo de información valiosa para que los apostadores futbolísticos puedan tomar las mejores decisiones frente a las casas de apuestas. Es por esto que en la interfaz web se presentan los marcadores de los últimos partidos jugados y de los partidos futuros.

En la siguiente imagen se muestra el resumen de resultados del análisis de

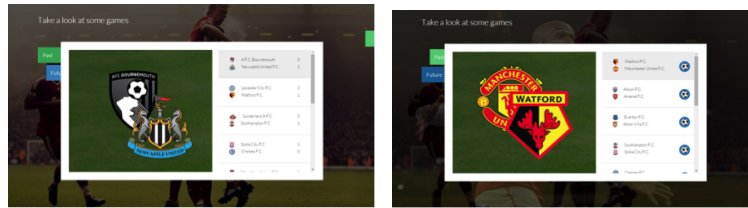


Figura 2: Datos partidos pasados (izquierda) y futuros (derecha)

[Check out what others are saying](#)



Figura 3: Twitter stream

datos y el modelo de predicción de betgram. Únicamente se presentan resultados probabilísticos para usuarios premium.

Como se mencionó anteriormente, la experiencia del usuario está acotada por su categoría (free o premium). La siguiente es la plataforma de suscripción, se pueden llevar a cabo suscripciones mensuales o suscripciones por temporada.

## Diseño de ingeniería

La materialización de *betgram* como producto, se da a través de una arquitectura robusta y escalable, para su correcto funcionamiento, *betgram* necesita poderosas herramientas para el análisis de grandes volúmenes de información, para la clasificación y análisis de lenguaje natural, y administración y almacenamiento de la información. Por esta razón, la base teórica necesaria para la producción de *betgram* está en términos clave como *Big data*, *Machine Learning*, *Data Mining*, *Social Network Analysis*, *Natural Language Processing*.

En términos muy generales, *betgram* necesita módulos encargados de las siguientes tareas.

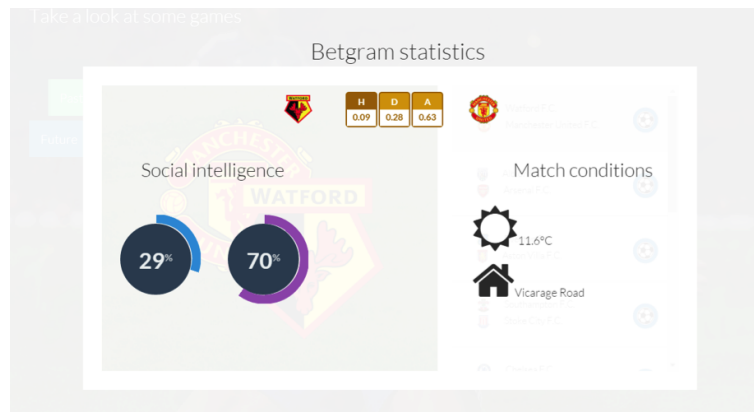


Figura 4: Resumen predictivo

- Recopilación de datos sobre fútbol en las redes sociales, especialmente twitter. Este componente, está encargado de recolectar estadísticas deportivas, información relevante y pasión por equipo en las redes sociales. Además, de esto, está encargada de administrar, analizar y servir la información relacionada. Este componente, se llama *betgram-tweet-engine*
- Recopilación de datos sobre fútbol en las noticias locales y de mayor importancia. Este componente, está encargado de recopilar estadísticas deportivas e información relevante sobre los partidos futuros en corto plazo. Además, de esto, está encargada de administrar, analizar y servir la información relacionada. Este componente, se llama *betgram-news-engine*
- Despliegue de la visualización de la interfaz web de *betgram*, comunicación dinámica con el servidor central de *betgram*. Esto es, la capa encargada de la comunicación de primer nivel con los usuarios y visitantes de *betgram*. Este componente, se llama *betgram-frontend*.
- Servir la aplicación web de manera robusta, disponer la información y estructurarla para entregarla al usuario final, disponer la parte dinámica de la aplicación web, disponer las garantías necesarias y administrar los datos de los usuarios. Este componente, se llama *betgram-backend*.

Por propósitos de rapidez en el desarrollo, la mayoría de capacidades de inteligencia artificial, serán provistas por librerías o servicios pertinentes, sin embargo, el equipo cuenta con el conocimiento necesario para realizar el modelo predictivo basado en máquinas de soporte vectorial, análisis de series de tiempo, y administración y análisis de grandes bancos de información estructurada y no estructurada. Para resumir de forma eficiente el despliegue y organización de *betgram*, se adjunta la figura 1.

La estrategia de análisis de la información relevante para la producción del producto final para el usuario, consiste en pesar de manera dinámica el aporte

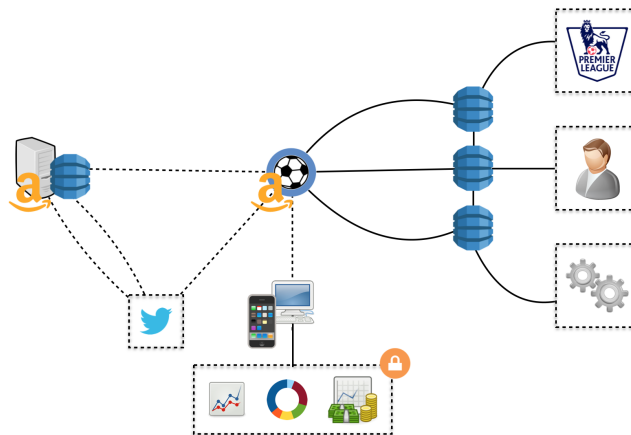


Figura 5: Arquitectura de *betgram*

de varias variables a la predicción final, este enfoque, con las variables típicas del análisis estadístico clásico, por último, incorporar el peso de las métricas sociales si la información contenida en ellas resulta relevante. Las construcciones teóricas clásicas, consisten en analizar la información con técnicas probabilísticas provenientes del aprendizaje supervisado, siendo las entradas finales del modelo variables tales como los goles marcados en la primera mitad del partido, el ganador, los indicadores deportivos de las casas de apuesta, el clima, la asistencia etc.

## Alcance del proyecto – Plan de implementación

*betgram* es un proyecto ambicioso con un amplio espectro de posibilidades de desarrollo y fine-tuning. Para efectos de PMC, acordamos con nuestros mentores acotar la complejidad del proyecto a partir de los siguientes lineamientos:

- Restringir el alcance en contenido de *betgram* a la liga profesional de fútbol de Inglaterra. La Barclays Premier League (BPL) hace parte de las cuatro grandes ligas europeas y es uno de los principales temas motivadores de producción de contenido deportivo en inglés. Este último hecho es particularmente relevante para *Betgram* puesto que una parte importante de las librerías disponibles para procesamiento de lenguaje natural están calibradas para el procesamiento de textos en inglés.
- Restringir el alcance en complejidad de *betgram* al cálculo de métricas simples derivadas del análisis del contenido extraído de Twitter. Estas métricas pueden incluir conteos de frecuencia relativa o análisis conven-

cional de sentimiento. En cualquier caso, las métricas deberán ser los suficientemente simples para ser calculadas utilizando los recursos humanos, de hardware y de red disponible en las semanas restantes.

Teniendo estas restricciones en cuenta, el equipo tiene como objetivo cumplir con el plan de implementación detallado en la figura 4.

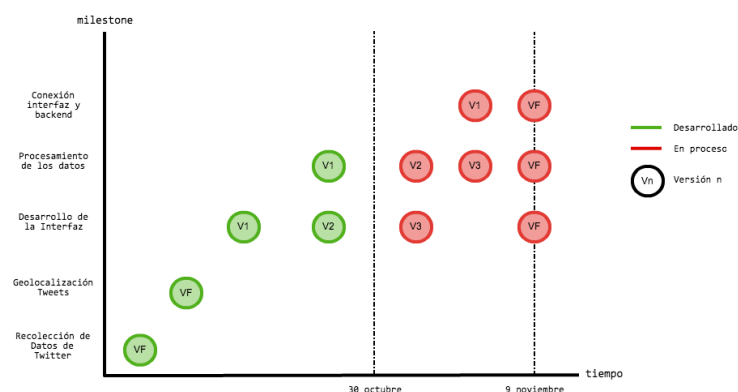


Figura 6: Milestones de *betgram*

## Plan de evaluación

*betgram* será evaluado en dos ejes fundamentales a su propuesta de valor. Por una parte, la naturaleza estocástica de los temas que trata *betgram* exige que sus modelos predictivos sean sujetos de validación estadística iterativa. Para fortuna de esta iniciativa, la BPL ofrece una oportunidad de ajustar el ejercicio predictivo por semana. Las fechas 12 y 13 del calendario de la BPL (7 y 14 de noviembre respectivamente) serán una primera oportunidad de realizar esta validación estadística.

Por otra, *betgram* es una iniciativa orientada a la prestación de un servicio y como tal debe tener la validación de sus la experiencia de sus usuarios como insumo fundamental para el mejoramiento continuo. De hecho, *betgram* nació justamente de ese intercambio de sensaciones entre el equipo de trabajo y los potenciales usuarios. Nos proponemos mantener esta cercanía con nuestros usuarios. Las redes sociales de la Universidad han sido fundamentales para este propósito. Una vez el prototipo sea funcional estaremos invitando a los usuarios que hemos contactado en el pasado a interactuar con el. Así mismo, esperamos que la existencia del prototipo nos permita atraer la atención de nuevos usuarios ajenos a las redes sociales de la Universidad.

## Organización del equipo

Para implementar, respaldo económico, y planes estratégicos de *betgram*, el equipo se ha dividido las labores más importantes para *betgram* de forma afín a cada uno, y de manera equilibrada, garantizando a su vez que cada miembro aprendiera y se sintiera cómodo con el desarrollo de las tareas asignadas. Se garantizó mediante la organización, que el desarrollo de los cuatro componentes fundamentales de *betgram* fueran desarrollados de manera incremental y simultánea, priorizando los requerimientos o servicios que tuvieran influencia sobre otros. A continuación, se listan las tareas principales de cada miembro del grupo.

- Juan Camilo Bages Prada: Implementación, diseño y desarrollo de los componentes de despliegue y visualización de la información principales de *betgram*, es decir, la interfaz web de *betgram*.
- Simón Ramírez Amaya: Diseño y desarrollo de los componentes fundamentales de *betgram* para la extracción de información no estructurada de las redes sociales y noticias. Análisis y administración dinámica de esta información. Además, diseño del modelo predictivo, corazón de *betgram*.
- Sebastián Camilo Valencia Calderón: Diseño de la arquitectura de solución, desarrollo de las interfaces de comunicación para los análisis de *betgram*, implementación de los modelos de predicción y clasificación relacionados con los partidos de fútbol.
- Meili Vanegas Hernández: Diseño de la arquitectura de solución, desarrollo de la parte relacionada a los modelos de fútbol, pertinentes para el desarrollo de los análisis posteriores, generación de documentación alusiva a *betgram* y diseño principal de *betgram-backend*.

## Primeros resultados

Los resultados preliminares de *betgram* son alentadores. Se entrenó un algoritmo de clasificación mediante redes bayesianas sobre los resultados observados de los últimos 20 años de enfrentamientos en la BPL. El algoritmo tomaba como input los odds históricos de las casas de apuesta y los combinaba con las métricas sociales extraídas del stream de tweets de *betgram*.

El modelo arroja un 54.35 % de acierto utilizando validación cruzada de 10 pliegues sobre la muestra de entrenamiento. Nuestro modelo, más allá de incorporar únicamente métricas sociales simples y de no haber sido evaluado de manera iterativa debido a las restricciones de tiempo, tienen un performance superior al estándar de la industria (50 %).



## Bibliografía

- *Asur, S. Humberman, B. (2010). Predicting the Future with Social Media. Social Computing Labs. HP Labs, Palo Alto California.*
- *Sinha, S. et al. Humberman, B. (2010). Predicting the NFL using Twitter. Department of Computer Science, Arkansas State University..*
- *Yu, S. Kak, S. (2013). A survey of prediction using social media. Department of Computer Science, Oklahoma State University..*
- *Snyder, J. Schapire, S. (2013). What Actually Wins Soccer Matches: Prediction of the 2011-2012 Premier League for Fun and Profit. Department of Computer Science, Washington University..*
- *Langseth, H. (2013). Beating the bookie: A look at statistical models for prediction of football matches. Department of Computer and Information Science, Norwegian University of Science and Technology..*