

**Novática**, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software).

<<http://www.ati.es/novatica/>>  
<<http://www.ati.es/reicis/>>

ATI es miembro fundador de **CEPIS** (Council of European Professional Informatics Societies) y es representante de España en **IFIP** (International Federation for Information Processing); tiene un acuerdo de colaboración con **ACM** (Association for Computing Machinery), así como acuerdos de vinculación o colaboración con **AdaSpain**, **AI2**, **ASTIC**, **RITS** e **HispanLinux**, junto a la que participa en **ProInnova**.

#### Consejo Editorial

Ignacio Aguiló Sousa, Guillem Alsina González, María José Escalona Cuarema, Rafael Fernández Calvo (presidente del Consejo), Jaime Fernández Martínez, Luis Fernández Sanz, Didac Lopez Viñas, Celestino Martín Alonso, José Oñofre Montes Andrés, Francesc Noguera Puig, Ignacio Pérez Martínez, Andrés Pérez Payeras, Viktu Pons i Colomer, Juan Carlos Vigo López

#### Coordinación Editorial

Llorenç Pagés Casas <pagues@ati.es>

#### Composición y autoedición

Jorge Lácer Gil de Ramales

#### Traducciones

Grupo de Lengua e Informática de ATI <<http://www.ati.es/gt/lengua-informatica/>>

#### Administración

Tomás Brunete, María José Fernández, Enric Camarero

#### Secciones Técnicas - Coordinadores

##### Acceso y recuperación de la información

José María Gómez Hidalgo (Optenet), <jmgomez@yahoo.es>

Manuel J. María López (Universidad de Huelva), <manuel.maria@dieisia.uhu.es>

##### Administración Pública electrónica

Francisco López Crespo (MAE), <flc@ati.es>

Sebastià Justicia Pérez (Diputación de Barcelona), <sjusticia@ati.es>

##### Arquitecturas

Enrique F. Torres Moreno (Universidad de Zaragoza), <enrique.torres@unizar.es>

José Filich Cardo (Universidad Politécnica de Valencia), <jfilich@disca.upv.es>

##### Auditoría SITIC

Marina Tourino Trolitfo, <marinatourino@marinatourino.com>

Sergio Gómez-Landero Pérez (Endesa), <sergio.gomezlandero@endesa.es>

##### Derecho y tecnologías

Isabel Hernández Collazos (Fac. Derecho de Donostia, UPV), <isabel.hernandez@ehu.es>

Elena Davara Fernández de Marcos (Davara & Davara), <edavara@davara.com>

##### Enseñanza Universitaria de la Informática

Cristóbal Parja Flores (DSIP-UCM), <cparja@sip.ucm.es>

J. Ángel Velázquez Iturbide (DLSI, URJC), <angel.velazquez@urjc.es>

##### Entorno digital personal

Andrés Marín López (Univ. Carlos III), <amarin@it.uc3m.es>

Diego Gachet Pérez (Universidad Europea de Madrid), <dgachet@uem.es>

##### Estandares Web

Encarna Quesada Ruiz (Virati), <encarna.quesada@virati.com>

José Carlos del Arco Prieto (TCP, Sistemas e Ingeniería), <jcarco@gmail.com>

##### Gestión del Conocimiento

Juan Baiget Solé (Cap Gemini Ernst & Young), <juan.baiget@ati.es>

##### Gobierno Cooperativo de las TI

Manuel Palao García-Suñto (ATI), <manuel@palao.com>

Miguel García-Mendoza (ATI), <mgarciamendoza@titrendsinstitute.org>

##### Informática y Filosofía

José Ángel Olivás Varela (Escuela Superior de Informática, UCLM), <joseangel.olivas@uclm.es>

Roberto Feltrero Oreja (UNED), <rfeltrero@gmail.com>

##### Informática Gráfica

Miguel Chover Sellés (Universitat Jaume I de Castellón), <mchover@lsi.uji.es>

Roberto Vivó Hernando (Eurographics, sección española), <rvido@dsic.upv.es>

##### Ingeniería del Software

Javier Dolado Cosin (DLSI-UPV), <dolado@lsi.ehu.es>

Daniel Rodríguez García (Universidad de Alcalá), <daniel.rodriguez@uah.es>

##### Inteligencia Artificial

Vicente Boti Navarro, Vicente Julián Inglada (DSIC-UPV), <(vbotti,vinglada)@dsic.upv.es>

##### Interacción Persona-Computador

Pedro M. Latorre Andrés (Universidad de Zaragoza, AIPO), <platorre@unizar.es>

Francisco L. Gutierrez Vela (Universidad de Granada, AIPO), <fgutierrez@ugr.es>

##### Lenguaje e Informática

M. del Carmen Ugarte García (ATI), <cugarte@ati.es>

##### Lenguajes Informáticos

Oscar Belmonte Fernández (Univ. Jaime I de Castellón), <belmonte@lsi.uji.es>

Inmaculada Coma Taty (Univ. de Valencia), <inmaculada.coma@uv.es>

##### Lingüística computacional

Xavier Gómez Guimovart (Univ. de Vigo), <xgg@uvigo.es>

Manuel Palomar (Univ. de Alicante), <mpalomar@lsi.ua.es>

##### Mundo estudiantil y jóvenes profesionales

Federico G. Mon Trotti (RITS), <gmon.trotti@gmail.com>

Mikel Sáizar Peña (Asoc. de Jóvenes Profesionales, Junta de ATI Madrid), <mikelbo\_uni@yahoo.es>

##### Profesión Informática

Rafael Fernández Calvo (ATI), <rfcalvo@ati.es>

Miguel Sarrías Girón (ATI), <miguel@sarrias.net>

##### Redes y servicios telemáticos

José Luis Marzo Lázaro (Univ. de Girona), <joseluis.marzo@udg.es>

Juan Carlos López López (UCLM), <juancarlosllopez@uclm.es>

##### Robótica

José Cortés Arenas (Sopra Group), <joscortes@gmail.com>

Juan González Gómez (Universidad CARLOS III), <juan@iearobotics.com>

##### Seguridad

Javier Areñio Bertolin (Univ. de Deusto), <jareñio@deusto.es>

Javier López Muñoz (ETSI Informática-UMA), <jlm@lcc.uma.es>

##### Sistemas de Tiempo Real

Alejandro Alonso Muñoz, Juan Antonio de la Puente Alfaro (DIT-UPM), <(alalmon, juente)@dit.upm.es>

##### Software Libre

Jesús M. González Barahona (GSYC - URJC), <jgb@gsyc.es>

Israel Herráiz Taberner (Universidad Politécnica de Madrid), <isra@herraz.org>

##### Tecnología de Objetos

Jesús García Molina (DLS-UM), <jmolina@um.es>

Gustavo Rossi (LIFIA-UNLP Argentina), <gustavo@sol.info.unlp.edu.ar>

##### Tecnologías para la Educación

Juan Manuel Dodero Beardo (UC3M), <dodero@inf.uc3m.es>

César Pablo Córcoles Briogio (UOC), <ccorcoles@uoc.edu>

##### Tecnologías y Empresa

Didac Lopez Viñas (Universitat de Girona), <didac.lopez@ati.es>

Alonso Álvarez García (TID), <aag@tid.es>

##### Tendencias tecnológicas

Gabriel Martí Fuentes (Interbits), <gabi@atinet.es>

Juan Carlos Vigo (ATI), <juancarlosvigo@atinet.es>

##### TIC y Turismo

Andrés Aguayo Maldonado, Antonio Guevara Plaza (Univ. de Málaga), <(aguayo, guevara)@lcc.uma.es>

Las opiniones expresadas por los autores son responsabilidad exclusiva de los mismos. **Novática** permite la reproducción, sin ánimo de lucro, de todos los artículos, a menos que lo impida la modalidad de © o copyright elegida por el autor, debiéndose en todo caso citar su procedencia y enviar a **Novática** un ejemplar de la publicación.

#### Coordinación Editorial, Redacción Central y Redacción ATI Madrid

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <[novatica@ati.es](mailto:novatica@ati.es)>

#### Composición, Edición y Redacción ATI Valencia

Av. del Reino de Valencia 29, 46005 Valencia

Tfno. 963740173 <[novatica\\_prod@ati.es](mailto:novatica_prod@ati.es)>

#### Administración y Redacción ATI Cataluña

Via Laietana 46, ppal. 1ª 08003 Barcelona

Tfno. 934125235; fax. 934127713 <[secregen@ati.es](mailto:secregen@ati.es)>

#### Redacción ATI Andalucía

<[secreand@ati.es](mailto:secreand@ati.es)>

#### Redacción ATI Galicia

<[secregal@ati.es](mailto:secregal@ati.es)>

#### Suscripción y Ventas

<[novatica.subscriptions@atinet.es](mailto:novatica.subscriptions@atinet.es)>

#### Publicidad

Plaza de España 6, 2ª planta, 28008 Madrid

Tfno. 91 4029391; fax. 91 3093685 <[novatica@ati.es](mailto:novatica@ati.es)>

Imprenta: Derra S.A. Juan de Austria 86, 08005 Barcelona.

Depósito legal: B 15.154-1975 - ISSN: 0211-2124; CODEN NOVACD

Portada: "Mineral, vegetal, animal" - Concha Arias Pérez / © ATI

Diseño: Fernando Agresta / © ATI 2003

## editorial

**El "caso Snowden" y la seguridad de las redes de telecomunicación en resumen** > 02

**Soporte al negocio y práctica profesional: El sueño del buen editor** > 03

*Llorenç Pagés Casas*

## noticias de IFIP

**IFIP TC6 Latin American Tutorials in Networking (LATIN 2013)** > 03

*Ramon Puigjaner Trepal*

## monografía

### Minería de procesos

*Editores invitados: Antonio Valle Salas y Anne Rozinat*

**Presentación. Una perspectiva sobre la minería de procesos** > 05

*Antonio Valle Salas, Anne Rozinat*

**Minería de procesos: La objetivación de la intuición en los procesos de toma de decisiones en los negocios, más transparentes gracias al análisis de los datos** > 07

*Anne Rozinat, Wil van der Aalst*

**Minería de procesos: Obtenga una radiografía de sus procesos de negocio** > 11

*Wil van der Aalst*

**El viaje del descubrimiento de procesos** > 20

*Josep Carmona Vargas*

**Posibilidades de uso de la minería de procesos en ITSM** > 24

*Antonio Valle Salas*

**Optimización dirigida por minería de procesos de un proceso de aprobación de préstamos al consumo** > 31

*Arjel Bautista, Lalit Wangikar, S.M. Kumail Akbar*

**Mejoramiento de procesos con técnicas de minería de procesos, simulación y optimización: Caso de estudio** > 41

*Santiago Aguirre Mayorga, Carlos Alberto Parra Rodríguez*

**Detección de cambios temporales en los procesos de negocio mediante el uso de técnicas de segmentación** > 49

*Daniela Lorena Luengo Mundaca, Marcos Sepúlveda Fernández*

## secciones técnicas

### Referencias autorizadas

 > 57

### visiones sobre Lenguajes de Programación

**Cómo la metáfora de objetos llegó a ser el principal paradigma de programación** > 62

*Jesús J. García Molina*

**Elección de lenguajes de programación para la enseñanza universitaria** > 67

*Baltasar García Pérez-Schofield*

**La importancia de la labor del programador. ¿Qué se espera? ¿Cómo se prepara?** > 70

**Análisis desde los lenguajes de programación**

*Óscar Belmonte Fernández, Carlos Granell Canut*

**Para pensar** > 79

*Rafael Martínez Martínez*

**Programando caminos y resolviendo necesidades** > 81

*Alejandro Fuentes Penna*

## sociedad de la información

### Programar es crear

**El problema del CUIT (corrección del publicado en el número anterior) (Competencia UTN-FRC 2012, problema D, enunciado)** > 82

*Julio Javier Castillo, Diego Javier Serrano, Marina Elizabeth Cárdenas*

## Asuntos Interiores

**Coordinación editorial / Programación de Novática / Socios Institucionales** > 83

Tema del próximo número:

**"Eficiencia energética en centros de proceso de datos"**

Arjel Bautista, Lalit Wangikar,  
Syed M. Kumail Akbar  
CKM Advisors, New York (EEUU)

<{abautista,lwangikar,sakbar}@ckmadvisors.com>

# Optimización dirigida por minería de procesos de un proceso de aprobación de préstamos al consumo

## 1. Introducción

A medida que el rol del Big Data se vuelve prevalente en esta era dirigida por la información [1][2][3], los negocios de todo el mundo buscan constantemente formas de aprovechar estos recursos potencialmente valiosos. El 2012 *Business Processing Intelligence Challenge* ("Desafío de Inteligencia de Procesamiento de Negocio", BPIC 2012) es un ejercicio de análisis de uno de esos conjuntos de datos utilizando una selección de herramientas comerciales, propietarias y de código abierto, y su combinación con percepciones creativas, para entender mejor el rol de la minería de procesos en el lugar de trabajo actual.

### 1.1. Enfoque y alcance

La situación representada en BPIC 2012 se enfoca en los procesos de aprobación de préstamos y descubiertos de una institución financiera real en los Países Bajos. En nuestro análisis de esta información hemos buscado entender en mayor detalle y a múltiples niveles de granularidad los procesos de negocio subyacentes. También hemos buscado identificar oportunidades para mejorar la eficiencia y efectividad del proceso global.

Específicamente, intentamos investigar en detalle las siguientes áreas:

- Desarrollar una comprensión minuciosa de los datos y del proceso subyacente.
- Entender las actividades y los puntos de decisión críticos.
- Mapear el ciclo de vida de una solicitud de préstamo desde el inicio hasta su resolución definitiva.
- Identificar diferencias en el desempeño a nivel de recursos y oportunidades para intervenciones en los procesos.

Siendo como somos recién llegados a la minería de procesos, en CKM Advisors, queríamos usar esta oportunidad para poner en práctica lo aprendido en esta disciplina. También intentamos combinar las herramientas de minería de procesos con métodos analíticos tradicionales, para construir un cuadro más completo. Estamos convencidos de que al adquirir experiencia nuestro enfoque se volverá más refinado y cada vez más conducido por métodos desarrollados específicamente para la minería de procesos.

Intentamos ser tan amplios como fuera posible en nuestro análisis y profundizar donde pudiésemos. Si bien hemos realizado un análisis

**Traducción:** Eduardo Rodríguez Ringach (TOGAF, CGEIT, CRISC)

**Resumen:** Un registro de eventos (262.200 eventos; 13.087 casos) del proceso de aprobación de préstamos y descubiertos de un banco en los Países Bajos fue analizado usando diversas técnicas analíticas. Mediante una combinación de enfoques basados en hojas de cálculo, capacidades de minería de procesos y analítica exploratoria, examinamos los datos en gran detalle y a múltiples niveles de granularidad. Presentamos nuestros hallazgos sobre como desarrollamos un profundo entendimiento del proceso, evaluamos áreas potenciales de mejora de eficiencia e identificamos oportunidades para realizar predicciones basadas en conocimiento sobre el resultado final de una solicitud de préstamo. También discutimos desafíos peculiares del trabajo con tales datos, y oportunidades para mejorar el impacto de dichos análisis mediante la incorporación de elementos de datos adicionales.

**Palabras clave:** Analítica de datos, Big Data, inteligencia de procesos de negocio, minería de procesos.

### Autores

**Arjel Bautista** es consultor de CKM Advisors, involucrado en el desarrollo de técnicas innovadoras de reingeniería de procesos y de investigación analítica en la empresa. En sus proyectos, ha desplegado una combinación de herramientas de minería de datos de última generación y el análisis estratégico tradicional para resolver diversos problemas relacionados con los procesos de negocio. También ha desarrollado estrategias para el análisis de texto no estructurado y otras fuentes de datos no tradicionales. Arjel ha obtenido su maestría y doctorado en química por la Universidad de Yale y un título de grado en Bioquímica por la UC San Diego.

**Lalit Wangikar** es socio de CKM Advisors. Como consultor, ha asesorado a clientes en especial en el sector de servicios financieros y las industrias de seguros y servicios de pago. Su área principal de especialización es el uso de *Big Data* y Analítica para conducir el impacto de negocio a través de todas las áreas empresariales clave como marketing, riesgo, operaciones y cumplimiento normativo. Ha trabajado con clientes en Norteamérica, Reino Unido, Singapur e India. Antes de asociarse a CKM Advisors, Lalit fue responsable de la práctica de Analítica de Decisiones para EXL Service / Inductis. Con anterioridad trabajó como consultor para Deloitte Consulting y Mitchell Madison Group, donde asesoró a clientes en los sectores verticales de banca y mercados de capitales.

**Syed M. Kumail Akbar** es consultor de CKM Advisors donde forma parte del equipo de Analítica, y colabora en minería de datos, mapeo de procesos y analítica predictiva. Previamente ha trabajado en proyectos de estrategia y operaciones en la industria de servicios financieros. Antes de unirse a CKM, trabajó como asistente de investigación en el Centro de Análisis Cuantitativo así como en el Departamento de Física de la Universidad Wesleyan. También fue cofundador de Possibilities Pakistan, una Organización No-Gubernamental dedicada a proveer acceso a orientación vocacional universitaria para los estudiantes de instituto en Pakistán. Syed se ha graduado en Física y Matemática-Economía por la Universidad de Wesleyan.

lisis detallado en unas pocas áreas, no hemos cubierto en nuestro análisis todas las posibles áreas de la minería de procesos. La determinación de las áreas que no hemos cubierto (por ejemplo, el análisis de redes sociales) se debe exclusivamente a nuestra propia comodidad y familiaridad con la materia en cuestión y no necesariamente por una limitación de los datos.

## 2. Materiales y métodos

### 2.1. Entendiendo los datos

Los datos capturan eventos de procesos para

13.087 solicitudes de préstamo/descubierto en un período de seis meses, entre octubre de 2011 y marzo de 2012. El registro de eventos está compuesto por un total de 262.200 eventos para estos casos, comenzando con un cliente que envía una solicitud y que culmina con la finalización de esa solicitud en un evento de aprobación, cancelación o rechazo. Cada solicitud tiene un único atributo, AMOUNT\_REQ, que indica la cantidad pedida por el solicitante. Para cada evento el extracto muestra el tipo de evento, la etapa del ciclo de vida (Programado, Iniciado, Finali-

Tipo	Descripción
“A_” Eventos de solicitud	Se refiere a estados de la propia solicitud. Cuando un cliente inicia una solicitud, personal del banco realiza el seguimiento para completar la solicitud si es necesario y para facilitar la toma de decisiones sobre las solicitudes.
“O_” Eventos de oferta	Se refiere a estados de una oferta comunicada al cliente.
“W_” Eventos de trabajo	Se refiere a estados de elementos de trabajo que ocurren durante el proceso de aprobación. Estos eventos capturan la mayoría del esfuerzo manual llevado a cabo por los recursos del banco durante el proceso de aprobación de la solicitud. Los eventos describen esfuerzos realizados durante diversas etapas del proceso de solicitud: <ul style="list-style-type: none"> <li>– <i>W_Afhandelen leads</i>: Seguimiento de presentaciones iniciales incompletas.</li> <li>– <i>W_Completeren aanvraag</i>: Compleción de solicitudes pre-aprobadas</li> <li>– <i>W_Nabellen offertes</i>: Seguimiento tras la transmisión de ofertas a solicitantes cualificados</li> <li>– <i>W_Valideren aanvraag</i>: Evaluación de la solicitud</li> <li>– <i>W_Nabellen incomplete dossiers</i>: Búsqueda de información adicional durante la fase de evaluación</li> <li>– <i>W_Beoordelen fraude</i>: Investigación de casos sospechosos de fraude</li> <li>– <i>W_Wijzigen contractgegevens</i>: Modificación de contratos aprobados</li> </ul>

Tabla 1. Nombres y descripciones de eventos.

zado), un indicador de recurso y el momento de finalización.

Los eventos describen pasos a lo largo del proceso de aprobación y son clasificados en tres tipos principales. La **tabla 1** muestra los tipos de eventos y nuestra comprensión de su significado

En sí mismo, el registro de eventos es una masa complicada de información de la cual resulta complicado inferir conclusiones lógicas. Por consiguiente, como otros investigadores han señalado [4][5], es menester someter al registro a cierto grado de pre-procesamiento para reducir su complejidad global, realizar conexiones visuales entre los pasos incluidos y auxiliar manualmente al análisis y la optimización de los conceptos de negocio. Aunque recibimos un registro de eventos rigurosamente pre-procesado que podía ser analizado sin inconvenientes usando herramientas de minería de procesos, procesamos todavía más los datos para construir extractos personalizados para diversos propósitos analíticos.

**2.2. Herramientas usadas para el análisis**

■ **Disco**: Adquirimos una versión de evaluación de Disco 1.0.0 (Fluxicon) y la utilizamos para exportar los datos a formatos adecuados para el análisis de hojas de cálculo. Disco fue especialmente útil para facilitar la visualización de flujos de proceso y excepciones típicas.

■ **Microsoft Excel**: Usamos Excel 2010 (Microsoft) para una exploración posterior más profunda en los datos pre-procesados. Excel ayudó especialmente para realizar funciones matemáticas básicas y avanzadas y para ordenar los datos, dos capacidades que destacan por su ausencia en la aplicación Disco.

■ **CART**: Usamos una versión de evaluación de la implementación de CART (Salford Systems) para llevar a cabo el análisis de segmentación preliminar de las solicitudes de préstamo con el fin de evaluar oportunidades de priorización del esfuerzo de trabajo.

**3. Entendiendo el proceso en detalle**

**3.1. Simplificación del registro de eventos**

Tras obtener el registro de eventos BPIC 2012, intentamos en primer lugar la reducción de su complejidad global identificando y removiendo eventos *redundantes*. Para los propósitos de este análisis, se considera que un evento es redundante si ocurre en forma concurrente con otro evento o a continuación de otro evento, de forma tal que el intervalo entre ambos sea mínimo (no más de dos segundos) respecto del marco temporal del caso completo.

El análisis inicial de los datos en crudo de *Disco* reveló un total de 4.366 variantes de orden de eventos entre los 13.087 casos representados. Conjeturamos que la remoción de hasta una sola secuencia de eventos

redundantes podía resultar en una reducción significativa del número de variantes. Esta simplificación se acentúa cuando el número de variantes removidas se multiplica por otras que ocurren más allá del evento inicial.

Además eliminamos dos tipos de eventos O-type (*O\_CANCELLED* and *O\_DECLINED*) que ocurren simultáneamente con *A\_CANCELLED* y *A\_DECLINED*, respectivamente.

No consideramos para su remoción eventos de tipo W-type ya que las fases de transición son cruciales para el cálculo de tiempo de trabajo dedicado a cada caso. Una vez removidos los eventos redundantes del registro de eventos, el número de variantes se redujo a 3.346, una mejora de casi el 25% respecto del conjunto de datos no filtrado.

Dicha consolidación puede ayudar en la simplificación de datos de proceso y facilitar un análisis más rápido. La complejidad de las variantes podría reducirse más si se entrevistase a los expertos de proceso del banco como ayuda para consolidar eventos que ocurren en forma conjunta y variantes secuenciales que no son críticas para el análisis del negocio.

**3.2. Determinación del flujo de caso estándar**

A continuación buscamos determinar el flujo de caso estándar para una solicitud exitosa contra el cual pudieran compararse todos los otros casos. Lo hicimos descargando el proyecto simplificado en Disco y filtrando todos los casos para el atributo *A\_APPROVED*. Fijamos a continuación al nivel más riguroso (0%) y los umbrales tanto de actividades como de caminos, lo que resultó en una representación idealizada del camino desde la presentación hasta la aprobación (ver **figura 1**).

**3.3. Entendiendo los resultados de las solicitudes**

Antes de embarcarnos en una revisión más detallada de los datos, entendimos que era necesario definir los resultados finales para todas las 13.087 solicitudes. Usando el flujo de caso estandarizado (ver **figura 1**) determinamos que todas las aplicaciones están sujetas a una de cuatro *suertes* en cada etapa del proceso de aprobación:

■ **Avance a la próxima etapa**: La solicitud avanza a la siguiente etapa del proceso.

■ **Aprobada**: Las solicitudes que son aprobadas y donde el cliente ha aceptado la oferta del banco son consideradas exitosas y son etiquetadas como Aprobadas, con el punto final representado por el evento *A\_APPROVED*.

■ **Cancelada**: La solicitud es cancelada por el banco o por pedido del cliente. Las solicitudes canceladas tienen un punto final de *A\_CANCELLED*.

“ Tras obtener el registro de eventos BPIC 2012, intentamos en primer lugar la reducción de su complejidad global identificando y removiendo eventos redundantes ”

Eventos Redundantes	Ocurrencia
A_PARTLYSUBMITTED	Inmediatamente a continuación de A_SUBMITTED en todos los 13.087 casos
O_SELECTED O_CREATED	Ambos en rápida sucesión antes de O_SENT para los 5.015 casos que fueron seleccionados para recibir ofertas. En algunos casos, O_CANCELLED (974 instancias), A_FINALIZED (2.907 instancias) o <i>W_Nabellen offeries-SCHEDULE</i> (1 instancia) ocurren entre O_SELECTED y O_CREATED en el proceso de creación de oferta.
O_ACCEPTED A_REGISTERED A_ACTIVATED	Ocurren los tres, en orden aleatorio, con A_APPROVED para las 2.246 solicitudes exitosas. En ciertos casos, O_ACCEPTED está intercalado entre estos eventos.

Tabla 2. Redundancias potenciales del registro de eventos.

■ **Denegada:** Luego de evaluar al solicitante se considera que este no es apto para recibir el préstamo o descubierto pedido. Las aplicaciones denegadas tienen un punto final de A\_DECLINED.

Nos aprovechamos del algoritmo de filtrado de Disco para definir un conjunto de posibles comportamientos de punto final. Se clasificó a 299 casos como *no* resueltos porque estaban en curso en el momento en que se capturaron los datos (es decir, no contenían puntos finales de A\_DECLINED, A\_CANCELLED o A\_APPROVED).

La **figura 2** muestra un flujo de proceso de alto nivel que marca como se disponen los casos en cada uno de los pasos de proceso clave. Este análisis nos ofrece percepciones útiles sobre el impacto general en el negocio de este proceso así como el flujo de caso global a través de los pasos de proceso críticos.

Observamos varias características de desempeño de línea base en la **figura 2**:

- ~26% de las solicitudes se rechazan instantáneamente (3.429 de las 13.087); lo que indica criterios estrictos de selección para mover una solicitud más allá del punto de partida.
- ~24% de las restantes (2.290 de las 9.658) se rechazan después del seguimiento inicial de

la oportunidad, indicando la existencia de un proceso continuo de selección de riesgo.

■ 754 de las 3.254 solicitudes que pasan a la etapa de validación (~23%) son rechazadas, indicando posibilidades para hacer más estricto el escrutinio directo en las etapas de solicitud u oferta.

**4. Evaluación del desempeño del proceso**

**4.1. Análisis a nivel de caso**

**4.1.1. Punto final de caso vs. duración general**

Para intentar evaluar cómo cambia la suerte de un caso particular respecto de la duración global, representamos gráficamente la relación entre estas dos variables y la presentamos superpuesta con el monto acumulado de tiempo de trabajo sobre la vida de estos casos. Excluimos los 3.429 casos que son rechazados instantáneamente al recibirse la presentación inicial de la solicitud porque no se gasta ningún esfuerzo en ellos. De este modo, nos propusimos visualizar el punto a partir del cual ejercer esfuerzo adicional rinde un retorno mínimo o nulo en forma de solicitudes completadas (cerradas).

La **figura 3** muestra una vista del ciclo de vida de todas las solicitudes, indizada en el momento de la presentación. Como se muestra en la figura, dentro de los primeros siete días las solicitudes siguen avanzando o son

rechazadas. En el Día 7, el número de casos aprobados comienza a elevarse, sugiriendo que este es el número mínimo de días requerido para completar los pasos en el flujo de caso estándar (ver **figura 1**). Las aprobaciones continúan hasta ~Día 23, cuando más del 80% de todos los casos que finalmente son aprobados han sido cerrados y registrados. Hay un salto significativo en el número de solicitudes canceladas en el Día 30, ya que se cancelan los casos inactivos donde no se ha recibido respuesta del solicitante después de empantanarse en las etapas de cuello de botella *Completeren aanvraag* o *Nabellen offeries*, probablemente siguiendo las políticas del banco.

Esto plantea la pregunta interesante de cuándo debería el banco detener los esfuerzos proactivos para convertir una solicitud en un préstamo, y si el banco debiera tratar a los clientes de distinta manera según los comportamientos que indican la verosimilitud de una aprobación final. Por ejemplo, el banco realiza un esfuerzo adicional de 380+ personas-día entre los Días 23 y 31, solo para terminar cancelando la mayoría de los casos pendientes al final del periodo. Con datos adicionales sobre la rentabilidad o el valor total durante la relación comercial con el cliente, se podría determinar un punto óptimo en el proceso a partir del cual no se obtendrá un valor positivo del esfuerzo adicional en los casos que no

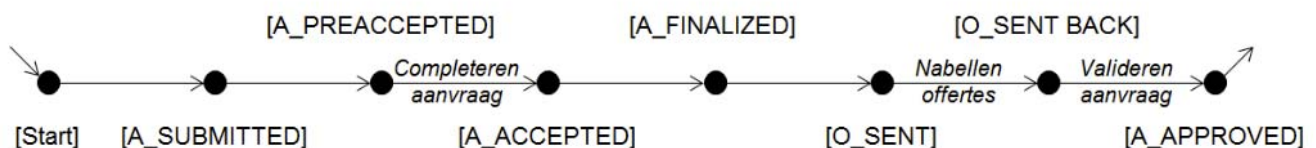


Figura 1. Flujo de Caso Estandarizado para las Solicitudes Aprobadas.

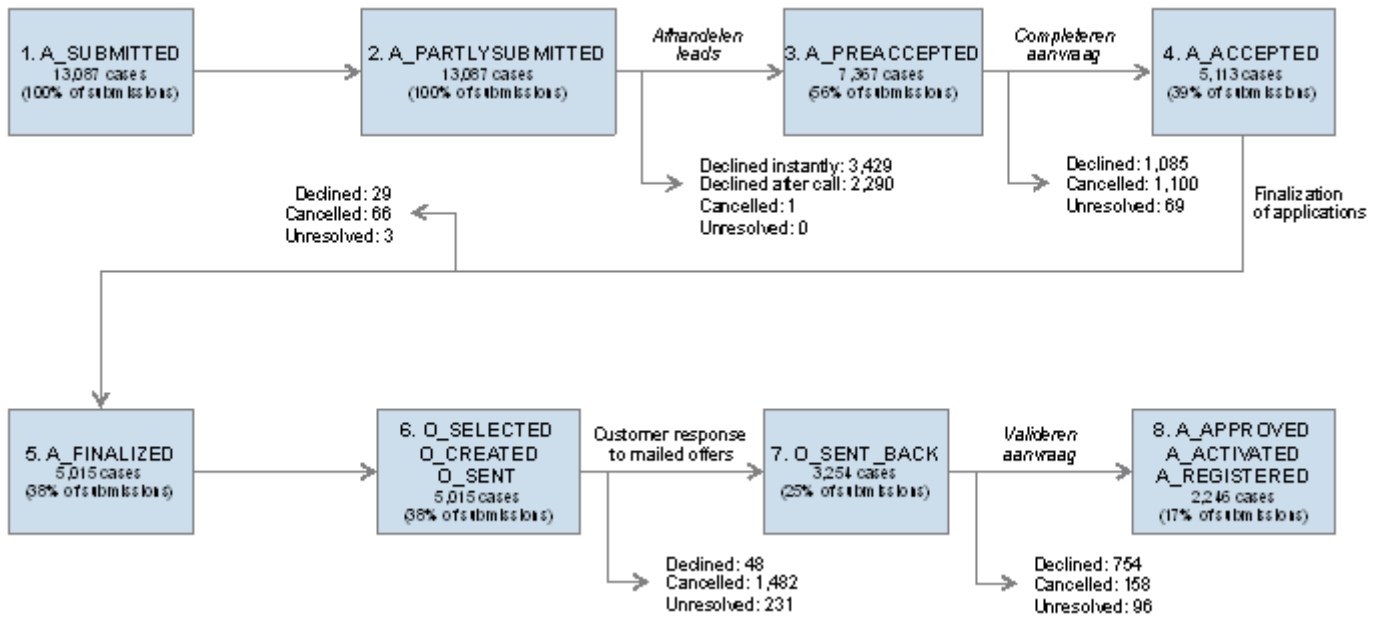


Figura 2. Pasos de proceso clave y flujo de volumen de solicitudes.

hayan alcanzado una cierta etapa.

**4.1.2. Segmentación de casos por importe solicitado**

Como cada caso está asociado con un im-

porte pedido por el solicitante, encontramos adecuado colocarlos en segmentos de aproximadamente el mismo número, ordenados por el valor total solicitado. En primer lugar quitamos los casos instantáneamente rechaza-

zados filtrándolos con Disco, ya que son resueltos en forma inmediata al ser presentados y no requieren ningún esfuerzo o pasos del proceso adicionales. Los 9.658 casos restantes (que incluye a los que están en curso)

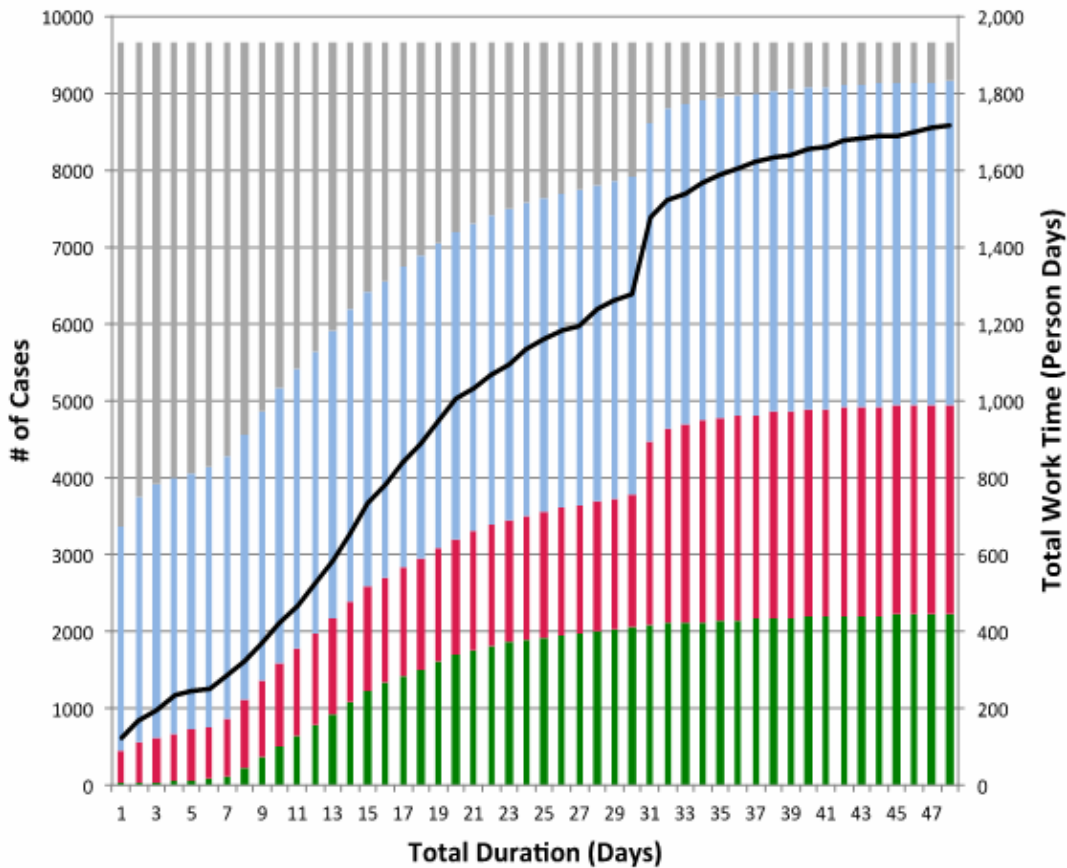


Figura 3. Distribución de casos por el resultado final y la duración, con esfuerzo de trabajo acumulado. Gris: Restantes en Curso, Azul: Rechazados Acumulados, Rojo: Cancelados Acumulados, Verde: Aprobados Acumulados. Excluye 3.472 Casos Instantáneamente Rechazados.

“ Nos propusimos visualizar el punto a partir del cual ejercer esfuerzo adicional rinde un retorno mínimo o nulo en forma de solicitudes completadas (cerradas)”

fueron luego divididos en deciles de unos 965-966 casos cada uno. Cada decil fue segmentado a su vez al clasificar los casos de acuerdo con su resultado final, y se examinaron las tendencias resultantes buscando correlaciones entre el porcentaje de aprobación con los montos pedidos (ver **figura 4**).

Inmediatamente observamos los porcentajes más altos de aprobación en los deciles 3 y 6, cuyos casos contenían rangos pedidos de 5.000-6.000 y 10.000-14.000, respectivamente. La razón exacta de este patrón no está clara; sin embargo especulamos que los solicitantes típicos eligen a menudo un número "redondo" en el que basar sus solicitudes (de hecho esto se refleja en los valores pedidos más frecuentes en el conjunto de datos: 5.000, 10.000 y 15.000). Quizás un cierto cambio de umbral de riesgo en el proceso de aprobación del banco cause un cambio abrupto en los porcentajes de aprobación.

**4.2. Análisis a nivel de evento**

**4.2.1. Cálculo de la duración del evento**

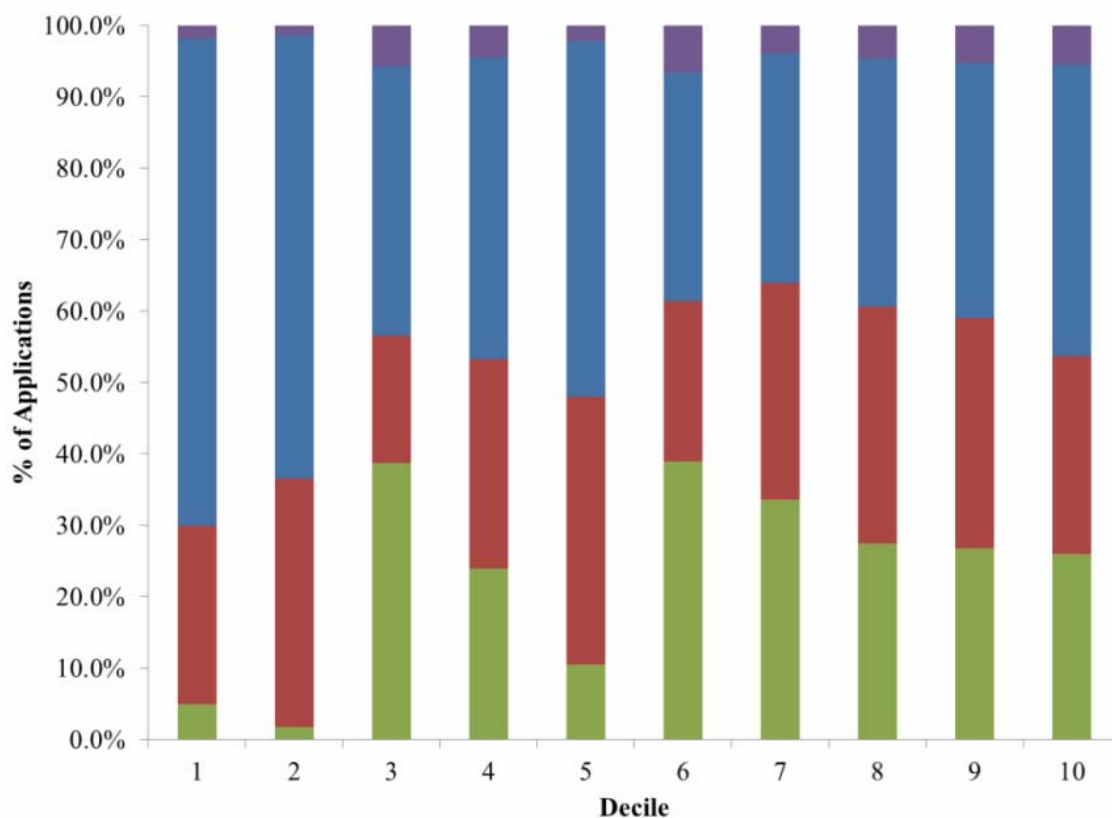
Buscamos conseguir una comprensión detallada de las actividades de trabajo embebidas en el proceso de aprobación, específicamente de aquellas que contribuyen con una cantidad significativa de tiempo o recursos a la resolución. El formato de los datos disponibles en este caso no se adecuaba fácilmente a este análisis. Usamos Excel para manipular los datos a nivel de evento tal como venían presentados y definimos el tiempo de trabajo (presumiblemente el esfuerzo real dedicado por recursos humanos) para cada evento como la duración de inicio a fin (transiciones START / COMPLETE, respectivamente). En cambio, el tiempo de espera fue definido como la latencia entre la programación del evento y el comienzo (SCHEDULE/START) o el tiempo transcurrido entre dos instancias de un único tipo de actividad así como entre el COMPLETE de un evento y el START de otro.

Como se muestra en la **tabla 3**, dos actividades *Completeren aanvraag* y *Nabellen Offertes*, contribuyen con un monto signifi-

cativo al tiempo total del caso según se lo representa en el registro de eventos. El tiempo de espera acumulado que se atribuye a cada uno de estos dos eventos puede llegar a ser tan alto como 30+ días por caso, ya que presumiblemente el banco realiza numerosos intentos de contactar al cliente antes de tener éxito. Mirando los datos con mayor detenimiento, nos dimos cuenta que el banco intenta contactar al cliente varias veces por día hasta el Día 30 para completar la solicitud, así como para el seguimiento de las ofertas que han sido realizadas pero que no han sido respondidas.

**4.2.2. Actividades iniciales vs. actividades de seguimiento**

El tiempo de trabajo promedio que se dedica a cada evento cambia si el banco lo realiza por primera vez o si se trata del seguimiento de un paso anterior en un caso particular (ver **figura 5**). Algunas diferencias en instancias iniciales y de seguimiento son mínimas (*Valideren aanvraag*), mientras que otras son más pronunciadas (*Beoordelen fraude*). En



**Figura 4.** Puntos finales de los casos (eje izquierdo), segmentados por importe pedido por el solicitante. Verde: Aprobado, Rojo: Cancelado, Azul: Rechazado, Violeta: En Curso.

“ Estos resultados sugieren que una oficina de especialistas realizando actividades únicas puede ser más adecuada que un ejército de recursos asignado con una miríada de tareas ”

	<i>Afhandelen Leads</i>	<i>Beoordelen Fraude</i>	<i>Completeren aanvraag</i>	<i>Nabellen Offertes</i>	<i>Nabellen Incomplete Dossiers</i>	<i>Valideren Aanvraag</i>
<b>Tiempo de Trabajo:</b>						
Aprobado	13.659	23	45.909	68.473	89.204	121.099
Cancelado	14.601	2	119.497	94.601	25.633	7.775
Rechazado	67.560	2.471	63.052	30.870	26.993	29.946
<b>Tiempo de Espera:</b>						
Aprobado	198.916	8.456	1.873.537	34.972.224	5.980.887	10.537.938
Cancelado	300.062	28.763	16.582.465	42.630.195	2.006.774	678.105
Rechazado	986.421	236.115	3.294.367	13.542.054	1.001.354	3.227.252

Tabla 3. Trabajo total y tiempo de espera por tipo de evento (en minutos).

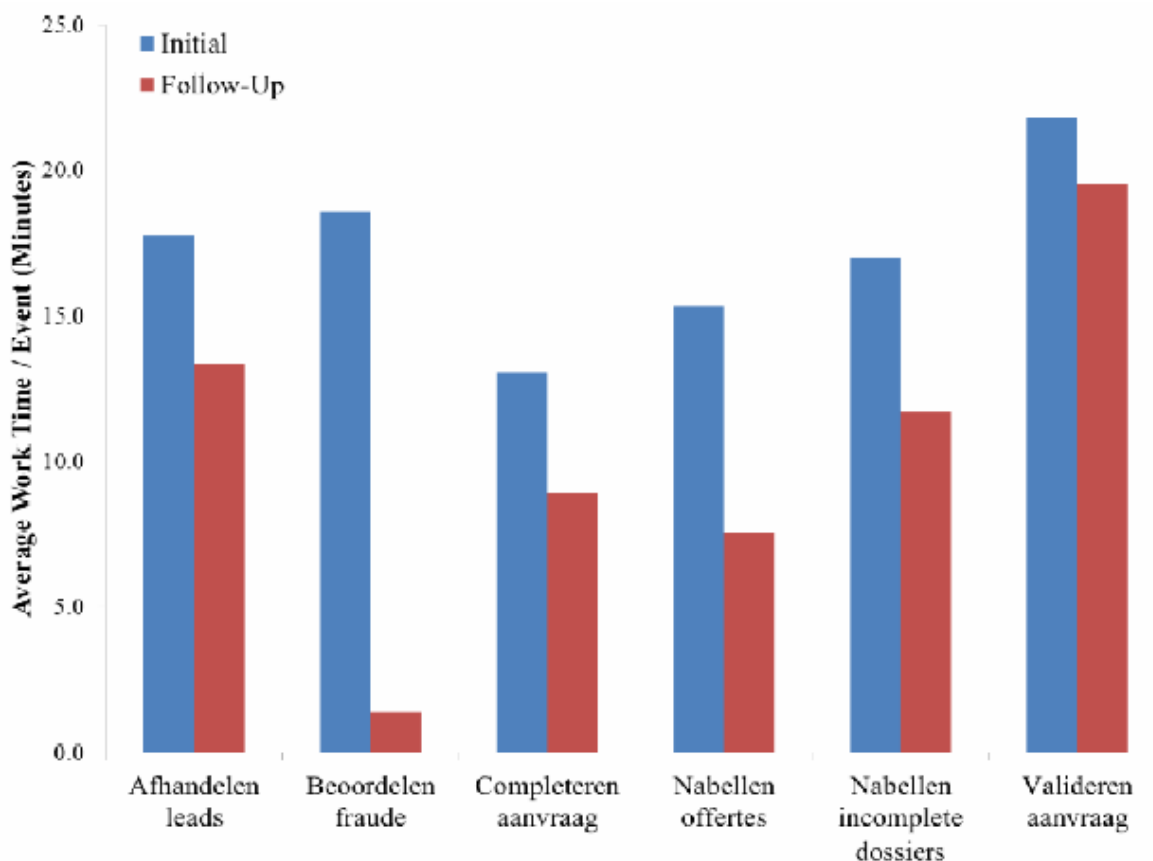


Figura 5. Comparación de tiempos de trabajo promedio, instancias de eventos iniciales vs. de seguimiento.

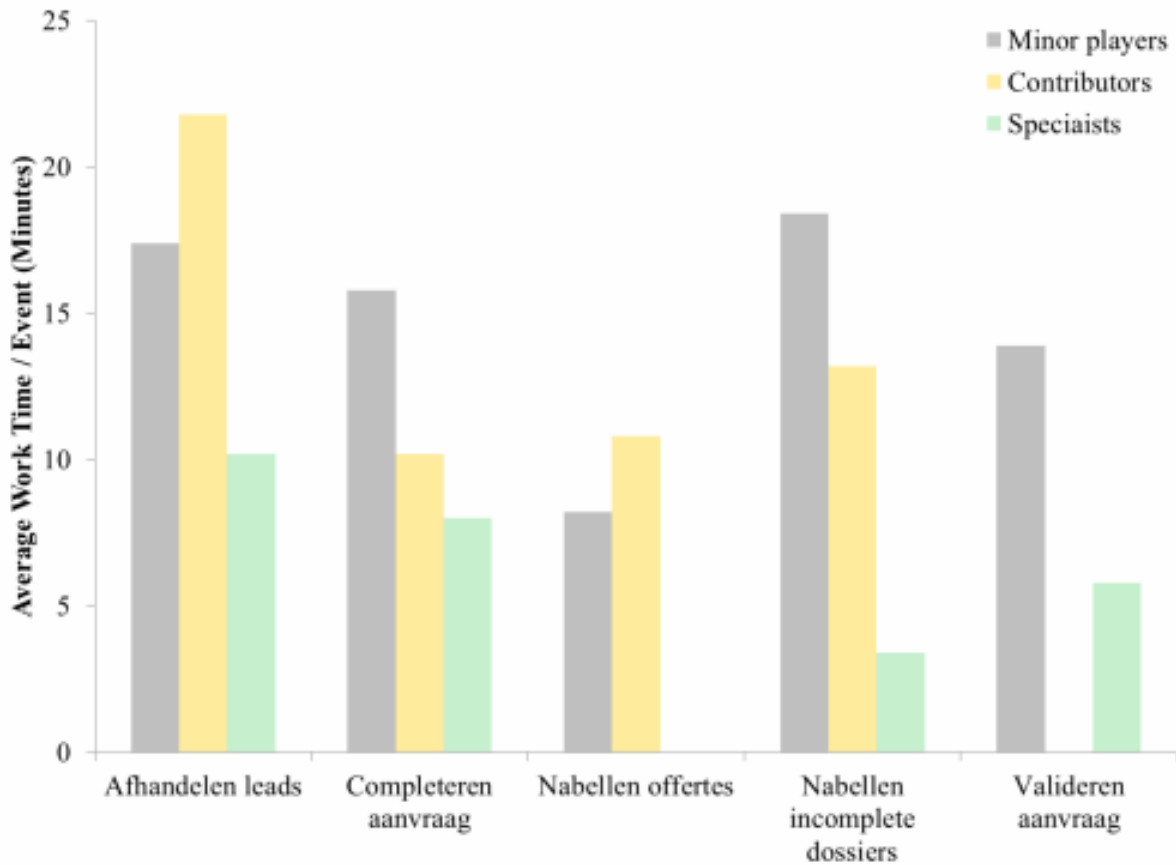


Figura 6. Tiempo de trabajo por evento, especialistas / contribuidores vs. colaboradores secundarios (*minor players*).

el caso de *Valideren aanvraag*, es probable que el banco sea lo más exhaustivo posible durante el proceso de validación, no importa cuántas veces antes haya visto una solicitud. Por otra parte, en la investigación de casos sospechosos de fraude, el banco puede haber llegado ya a una conclusión preliminar respecto de la solicitud y utilizar la instancia de seguimiento meramente para justificar su decisión.

Las instancias de seguimiento para aquellos eventos donde el banco debe contactar al solicitante a menudo tienen tiempos de trabajo promedio menores que sus contrapartes iniciales ya que estas actividades son las que tienen mayores posibilidades de quedar atrapadas en ciclos repetitivos, quizás debido a clientes que no responden. Uno puede aprovechar tales datos de eventos para entender el comportamiento del cliente y evaluar su utilidad potencial para la priorización del trabajo.

**4.3. Análisis a nivel de recursos humanos**

**4.3.1. Actividades de trabajo dirigidas por especialistas vs. generalistas**

Hemos construido el perfil de 48 personas (recursos humanos) que gestionaron al menos 100 eventos totales (excluyendo el recurso 112, ya que esta persona no gestiona otros eventos de trabajo que no sean la programa-

ción) y computamos el volumen de trabajo por el número de eventos gestionados por cada uno. Observamos nueve personas que invirtieron >50% de su esfuerzo en *Valideren aanvraag*, y un grupo destacado que realizó mayormente actividades de *Completeren aanvraag*, *Nabellen offertes* y *Nabellen incomplete dossiers*. Al parecer la validación la realiza un grupo dedicado de especialistas enfocado en este tipo de trabajo, mientras que las actividades de cara al cliente como *Completeren aanvraag*, *Nabellen offertes* y *Nabellen incomplete dossiers* podrían necesitar habilidades similares que son realizadas por otro grupo especializado.

A continuación examinamos el desempeño de recursos identificados como especialistas (>50% de los eventos de trabajo de un solo tipo) o contribuidores (25-50%) y los comparamos con aquellos que juegan solo un rol menor en actividades similares. A este fin tomamos el tiempo de trabajo total acumulado en una actividad por recursos pertenecientes a una categoría particular y calculamos promedios basados en el número total de eventos de trabajo realizados en esa categoría. Dos actividades, *Nabellen offertes* y *Valideren aanvraag*, no contienen especialistas o contribuidores, respectivamente y por lo tanto estas categorías fueron omitidas en las comparaciones para estas actividades.

Como se muestra en la **figura 6**, los especialistas dedicaron menos tiempo por instancia de evento que sus contrapartes, en algunos casos realizando las tareas hasta un 80% más eficientemente que los colaboradores secundarios. El desempeño de los contribuidores es mucho menos consistente, sin embargo, exhibiendo tiempos de trabajo promedio por caso que son más altos (*Afhandelen leads*, *Nabellen offertes*) o más bajos (*Completeren aanvraag*, *Nabellen incomplete dossiers*) que los de los colaboradores secundarios. Estos resultados sugieren que una oficina de especialistas realizando actividades únicas puede ser más adecuada para gestionar un número de casos más alto que un ejército de recursos asignado con una miríada de tareas.

**4.4. Aprovechamiento de datos de comportamiento para priorizar el esfuerzo de trabajo**

Uno de los objetivos de la minería de procesos es identificar oportunidades para dirigir la efectividad del proceso; esto es, lograr mejores resultados de negocio por el mismo o menor esfuerzo en un periodo igual o más corto. En particular, buscamos utilizar datos de eventos de proceso recogidos en una solicitud para priorizar mejor los esfuerzos de trabajo. Específicamente, nos propusimos entender si esto podría



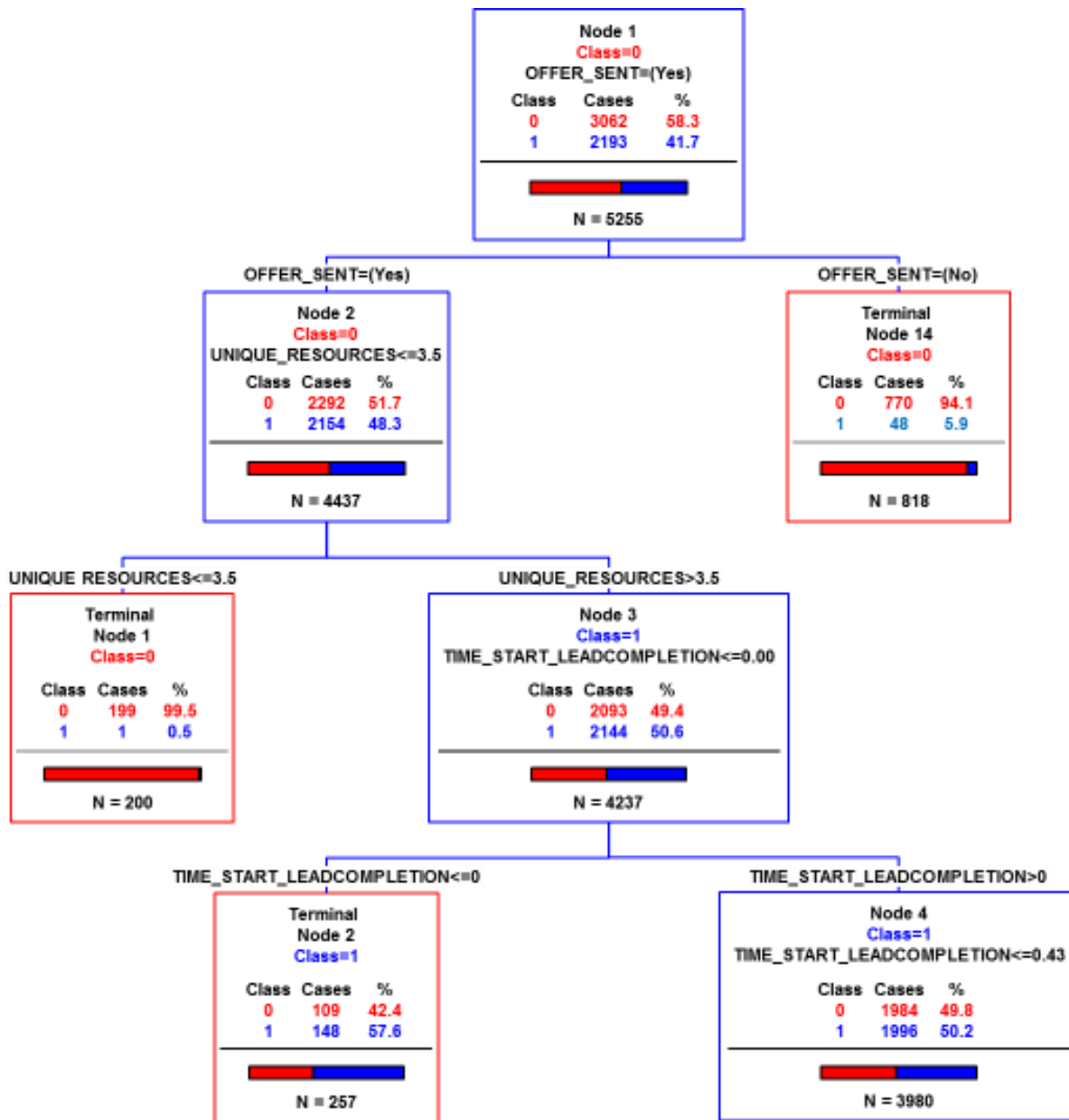


Figura 7. Vista parcial, árbol de segmentación CART.

hacerse en el quinto día a contar desde la presentación de la solicitud.

Con este fin, creamos un conjunto de datos a nivel de solicitud para 5.255 casos que duraron más de 4 días y donde el resultado final es conocido. Para estas solicitudes, capturamos todos los eventos desde la presentación hasta el final del día 4 y los usamos para calcular lo siguiente:

- Qué etapa había alcanzado la solicitud, y si había sido completada.
- Cuánto esfuerzo se había invertido en la solicitud.
- Cuántos eventos de cada clase ya habían sido registrados.
- Si la solicitud requería seguimiento de la oportunidad.

Intentamos encontrar segmentos clave en esta población con alta probabilidad de ser aprobados y aceptados, o bien con alta probabilidad de ser cancelados o rechazados. Lo hicimos segmentando los datos usando la técnica del Árbol de Clasificación y Regresión (*Classification and Regression Tree*, ver **figura 7**).

El árbol parcial anterior muestra dos segmentos con menos del 6% de tasas de aprobación. Los nodos terminales 1 y 14, con un total de 1.018 casos y con solamente 49 aprobaciones finales. El nodo 14, compuesto por 818 casos, muestra solicitudes incompletas donde el banco no pudo preparar ofertas para los clientes al final del Día 4. Tales solicitudes "lentas" tienen una probabilidad

menor del 6% de ser aprobadas, comparadas con un promedio del 42% para el grupo total de 5.255. El nodo 1 tiene solicitudes que son tocadas por 3 o menos recursos, uno de los cuales es el 112. Esto podría ser otro indicador de una solicitud lenta. Tales solicitudes carecen prácticamente de posibilidades de ser finalmente aprobadas.

Se podría repetir este análisis en diferentes etapas del ciclo de vida de la solicitud para colaborar con la priorización de esfuerzo. Este análisis preliminar indica un potencial significativo para reducir el esfuerzo en casos que podrían no llegar al estado final deseado. Análisis posteriores con demografía de clientes, detalles de la solicitud e información adicional acerca de los recur-

	<i>Afhandelen Leads</i>	<i>Completer en aanvraag</i>	<i>Nabellen Offertes</i>	<i>Nabellen Incomplete Dossiers</i>	<i>Valideren Aanvraag</i>
Tiempo de Trabajo Total (Minutos)	88.905	205.588	133.768	171.668	158.566
# Total de Tareas	5.041	20.830	10.426	19.748	7.819
Eficiencia Promedio del Especialista (Minutos / Evento)	10,2	8,0	10,8*	3,4	5,8
Tiempo Total de Trabajo Bajo la Eficiencia Promedio del Especialista (Minutos)	51.418	166.640	112.600	67.143	45.350
<b>Ahorros de Tiempo Proyectoados (Minutos)</b>	<b>37.487</b>	<b>38.949</b>	<b>21.167</b>	<b>104.525</b>	<b>113.216</b>

**Tabla 4.** Ahorros potenciales de tiempo asociados con la conversión de los generalistas actuales en especialistas de una única actividad. \*Ninguna de las personas que ejecutan *Nabellen offertes* fue identificada como especialista; por lo tanto se utilizó en su lugar a

los que trabajaron en tales casos podrían ayudar a refinar los hallazgos y sugerir pasos de acción específicos para mejorar la efectividad del proceso.

**5. Discusión**

**5.1. Tratamiento de los desafíos que presentan los datos**

**5.1.1. Gestión de la complejidad de los eventos**

La optimización del proceso de aprobación de préstamos es un ejercicio de racionalización de cada paso de la operación de principio a fin. Un punto notable que genera desafíos en la construcción de una visión del proceso racionalizada con herramientas de minería de procesos automatizadas es la cantidad y complejidad de los datos capturados. Si tales datos no se acompañan con un apropiado juicio de negocio, es posible perderse en la complejidad aparente (existen más de 4.000 variantes para un proceso que tiene 6-7 pasos clave). Hemos ilustrado este punto más arriba en nuestra discusión sobre eventos redundantes.

Recomendamos que esas complejidades sean abordadas en el momento del análisis, usando conocimiento del proceso y buen juicio empresarial, realizando pasos adicionales de pre-procesamiento de datos.

También es crítico escudriñar los datos de eventos por adelantado para entender todas las peculiaridades y construir formas de afrontarlas. Por ejemplo, una comparación del número de transiciones START y COMPLETE para eventos de tipo W-type en el conjunto de datos revela que hay 1.037 más transiciones COMPLETE que transiciones START. Como la marca temporal para estos eventos es única respecto de otras en el mismo *Case ID*, tienen el potencial de confundir mucho la acumulación de tiempo de trabajo y de espera para un caso particular y para recursos en la institución. Los designamos como errores de los sistemas y trabajamos con la primera transición COMPLETE que sigue a una START como la "correcta" para un tipo de evento de trabajo dado. En un proyecto real, validaríamos nuestra asunción revisando más en profundidad cómo surgen dichas instancias en el sistema y utilizando nuestra comprensión para tratar estas observaciones de forma correcta en nuestro análisis.

Como describimos en la **sección 3.1**, el registro de eventos también se beneficiaría de la consolidación de eventos que suceden concurrentemente, como los que ocurren cuando se aprueban solicitudes exitosas

(A\_APPROVED, A\_REGISTERED y A\_ACTIVATED). Esto no solo reduciría el tamaño general del archivo (que pasa a ser importante cuando crece el volumen de los datos) sino que también reduce la complejidad del registro inicial.

**5.2. Beneficios potenciales del re-despliegue de recursos**

**5.2.1. Redefinición de generalistas como especialistas**

Como mencionamos antes, las tareas que están implicadas en el proceso de aprobación de préstamos son realizadas por una mezcla de especialistas y generalistas. A lo largo de nuestro análisis concluimos que el banco podría beneficiarse de la especialización del trabajo, de forma que los recursos actuales se reasignen a puestos únicos para maximizar la eficiencia. En la **tabla 4** mostramos las ganancias potenciales que pueden obtenerse mediante dicha reestructuración. Si el banco puede mejorar el desempeño de todos quienes ejecutan una tarea al mismo nivel que los especialistas, estimamos un ahorro sustantivo y general de tiempo.

También evaluamos los ahorros potenciales asociados con el ajuste a la baja del grupo general de personas asignadas a estas tareas. Utilizando la cantidad promedio de tiempo de trabajo para los recursos que manejan más de 100 eventos totales (aproximadamente 16.000 minutos; de nuevo excluyendo el recurso 112), estimamos la oportunidad de reducir el esfuerzo de trabajo en un 35%.

**5.3. El poderío de la información adicional**

**5.3.1. Atributos adicionales a nivel de caso**

En su formato bruto, el registro de eventos BPIC 2012 es una mina de oro de información que, una vez decodificada, ofrece una vista detallada de un proceso de aprobación de préstamos al consumo. Sin embargo, esta información se fortalecería enormemente si se añadieran unos pocos puntos de datos clave. Como cada caso trae consigo un único atributo (el importe pedido por el solicitante) no tenemos forma de saber por qué ciertos casos son aprobados mientras que otros con montos pedidos y caminos idénticos son rechazados. Por consiguiente, sería útil conocer la demografía de los clientes, las relaciones actuales o anteriores con los clientes y detalles adicionales sobre el personal que ejecuta estos procesos. Con esta información podríamos construir recomendaciones específicas para cambiar el proceso y estimar con mayor precisión los beneficios probables de tales cambios.

**5.3.2. Rentabilidad del cliente y costes operativos**

Un conjunto final de datos que está notoriamente ausente del registro BPIC 2012 provis-

to es el coste global asociado con el proceso de aprobación de préstamos y el valor de cada solicitud de préstamo para el banco. Merecería la pena entender cuánto cuesta operar cada recurso y si este coste varía en base a las actividades que el personal realiza o el número de eventos en los que participa. Esta información también nos permitiría calcular un coste de adquisición promedio para cada solicitante, y consiguientemente entender el umbral mínimo por debajo del cual no tiene sentido económico aprobar una petición de préstamo entrante.

### 6. Conclusiones

A través del análisis comprehensivo del registro de eventos de BPIC 2012, hemos convertido un conjunto de datos bastante complejo en un flujo de trabajo de extremo a extremos claramente interpretable para un proceso de aprobaciones de préstamos y descubiertos. Examinamos los datos en múltiples niveles de granularidad, descubriendo percepciones interesantes en todos los niveles. A través de nuestro trabajo sacamos a la luz mejoras potenciales en varias áreas, incluyendo la revisión de procesos automatizados, la reestructuración de recursos claves y la evaluación de los procedimientos actuales de tratamiento de casos. De hecho, el análisis futuro será ayudado en buena medida por la inclusión de datos adicionales tales como información del cliente, políticas de gobierno, costes operativos y valor relativo del cliente.

Como parte de nuestro análisis, realizamos un ejercicio predictivo rudimentario mediante el cual determinamos el estado actual de casos en diversos días a lo largo del proceso de aprobación y cuantificamos sus posibilidades de aprobación, cancelación o rechazo. Esto nos permitió estimar la suerte de un caso basado en su desempeño y adaptar el proceso general para minimizar el estancamiento en cuellos de botella tradicionales en los casos. Aunque sea preliminar por su naturaleza, esto abre la puerta para futuros ejercicios de modelado más elaborado, quizás dirigidos por algoritmos informáticos sofisticados.

Si bien cubrimos varias áreas en este ejercicio, hay otras donde no realizamos un análisis detallado. El banco se beneficiaría de manera significativa al explorar esas áreas adicionales, como por ejemplo el análisis de redes sociales.

En conclusión, los procedimientos resaltados por BPIC 2012 destacan el rol e importancia de la minería de procesos en el lugar de trabajo moderno. Pasos que con anterioridad solo podían ser elucidados después de años de práctica y observación pueden ahora ser examinados utilizando una muestra de datos existentes. A medida que la era del *Big Data* continúa su marcha hacia el mundo del nego-

cio, pronosticamos que la minería de procesos será un elemento central en el cambio que convertirá preguntas en soluciones y problemas en beneficio sostenible.

### Reconocimientos

Agradecemos a la institución financiera que ha puesto a disposición estos datos para su estudio. Un agradecimiento especial a Fluxicon por facilitarnos una copia de evaluación de Disco acompañada de una copia del conjunto de datos BPIC 2012. También damos las gracias a Tom Metzger, Nicholas Hartman, Rolf Thrane y Pierre Buhler por sus útiles discusiones y percepciones. Nuestro agradecimiento se dirige también a Salford Systems, que puso a nuestra disposición su software en una versión de demostración.

### Referencias

- [1] W. Van der Aalst, A. Adriansyah, A.K. Alves de Medeiros, F. Arcieri, T. Baier *et al.* Process Mining Manifesto. *Business Process Management Workshops 2011, Lecture Notes in Business Information Processing*, vol. 99. Springer-Verlag, 2011.
- [2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011. <[http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)>.
- [3] R. Adduci, D. Blue, G. Chiarello, J. Chickering, D. Mavroyiannis *et al.* *Big Data: Big Opportunities to Create Business Value*. Technical report, Information Intelligence Group, EMC Corporation, 2011.
- [4] R.P.J.C. Bose, W.M.P. van der Aalst. Analysis of Patient Treatment Procedures: The BPI Challenge Case Study. *First International Business Process Intelligence Challenge*, 2011. <<http://bpmcenter.org/wp-content/uploads/reports/2011/BPM-11-18.pdf>>.
- [5] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011. ISBN-10: 3642193447.