

Trabajo de grado en modalidad de aplicación

## Desarrollo de una metodología de monitoreo predictivo de procesos en un sistema de manufactura auto-organizado

Laura María López Castro<sup>a,c</sup>, Sonia Geraldine Martínez Mayorga<sup>a,c</sup>, Nestor Eduardo Rodriguez Rubio<sup>a,c</sup>, Luna Violeta Lovera Abella<sup>a,c</sup>

Hugo Santiago Aguirre Mayorga<sup>b,c</sup>, Jose Fernando Jimenez<sup>b,c</sup>

<sup>a</sup>Estudiante de Ingeniería Industrial

<sup>b</sup>Profesor, Director del Proyecto de Grado, Departamento de Ingeniería Industrial

<sup>c</sup>Pontificia Universidad Javeriana, Bogotá, Colombia

---

### Resumen de diseño en Ingeniería

The objective of this project is to develop a methodology that allows the introduction of a predictive model in a flexible self-organized manufacturing system, allowing the system to make better decisions. Taking into consideration current competitiveness in productive markets and, the higher requirements in manufactured products, the development of improvements of various types of production systems is necessary. Additionally, taking opportune and accurate decision in ongoing processes allows to generate better results in various indicators such as the execution time of process and an adequate response to possible perturbations. This allows to improve the efficiency of the system. On the other hand, data analysis and industrial engineering tools through process mining, will allow to develop a methodology that permits the implementation of a predictive model of processes in a simulated flexible self-organized manufacturing system based on the manufacturing cell AIP - PRIMECA located in Valenciennes Université (France). Process mining tools such as *Apromore*, *Nirdizati*, ProM will be used as a basis for the development of the methodology and its implementation in the simulated system. It is expected that with the implementation of the methodology, the system will be more efficient.

*Key words: Flexible manufacturing system, self-organized system, process mining, predictive process monitoring, Netlogo, Nirdizati, Celonis, Apromore.*

---

### 1. Justificación y planteamiento del problema

Un sistema de manufactura flexible, es un sistema totalmente integrado, con características de gran escala de producción consistente en varios componentes de hardware y software que trabajan de manera cooperada (Tolio, 2009). Este tipo de sistemas tiene gran aplicabilidad en los ambientes productivos actuales, como en el relacionado con la celda de manufactura AIP - PRIMECA (Trentesaux et al, 2013), ubicada en la Universidad de Valenciennes en Francia.

Un sistema autoorganizado es aquel que tiene la capacidad de definir la secuencia de operaciones a ejecutar para cada producto, el agente que las ejecuta, y la ruta de cada producto hasta que finaliza su secuencia. La

aplicabilidad de la minería de procesos en el análisis de sistemas auto-organizados de manufactura, ya ha sido verificada anteriormente utilizando algoritmos como el denominado alpha, (un algoritmo de descubrimiento, que a partir de un registro de eventos se espera que modele con una red de petri el comportamiento real de un proceso) (Van Der Aalst, W. , 2016) que fue usado para comparar, por medio de una simulación, el desempeño de un sistema multiagentes en dos diferentes escenarios en un sistema de manufactura flexible auto-organizado de la Universidad de Valenciennes (Jimenez, Zambrano, Aguirre S & Trentesaux, 2018).

A partir del simulador de un sistema de manufactura flexible auto-organizado, se desarrolla una metodología basada en herramientas de la minería de procesos llegando a un monitoreo predictivo, que facilita la toma de decisiones del sistema mencionado, mejorando el control de su desempeño. El monitoreo predictivo de procesos permite mejorar la toma de decisiones a partir del control de los procesos en marcha, mientras que, un monitoreo de procesos tradicional sólo provee información sobre tiempos de ejecución de productos, recursos utilizados y ciertos indicadores para eventos particulares al final de una ejecución (Teinmaa, Dumas, Rosa & Maggi, 2017); el monitoreo predictivo de procesos detecta posibles problemas durante la ejecución de un proceso, permitiendo una toma de decisiones oportuna para mitigar consecuencias que se puedan presentar (Jalonen y Lönnqvist, 2009).

La consecución de dicha metodología, permite tanto un avance en cuanto al control desempeño de los sistemas auto-organizados de manufactura flexible, puesto que perturbaciones en el sistema de manufactura podrán ser afrontadas de una manera más eficiente, como la toma de decisiones para afrontar posibles retrasos en los pedidos de entrega (Verenich et al, 2018). El control de procesos en marcha, permite al mismo sistema de manufactura o en su defecto a quien controla el proceso, el ajuste de actividades, para un mejor aprovechamiento de los recursos disponibles.

Existen herramientas para la predicción de código abierto basadas en la minería de procesos tales como *Apromore* y *Nirdizati* desarrolladas por las universidades de Tartu y de Tecnología de Queensland (Australia). La primera de ellas es una plataforma web avanzada de análisis de procesos por la comunidad de administración de procesos de negocio, mientras la segunda es una herramienta que contiene diversas técnicas predictivas (Verenich, et al 2018). Algunas funcionalidades *Nirdizati*, fueron incluidas en *Apromore* (Verenich, et al 2018), por lo que hace más viable su uso para lo propuesto en el presente proyecto.

Combinar este tipo de herramientas, y su aplicación dentro de un contexto de sistemas auto-organizados de manufactura, contribuyendo a su desarrollo, será la propuesta presentada en el presente documento. El presente proyecto se tomó como profundización al tema de minería de procesos aplicado a una simulación de un sistema de manufactura (Jiménez et al, 2018), añadiendo elementos predictivos al proceso.

## **2. Antecedentes**

### **Contexto general**

Actualmente, las organizaciones en general, crean y adquieren una cantidad extensa de información constante de datos para cualquier tipo de proceso que se requiera para la producción, desarrollo y estructuración de algún bien o servicio. Adicional a lo anterior, la constante intervención de la globalización, y los avances tecnológicos a las industrias y organizaciones, las cuales otorgan ventajas competitivas, han dejado como consecuencia un aumento de gran impacto en la adquisición de datos. Dichos datos obtenidos no se procesan de manera manual, sino que, por lo contrario, se almacenan de manera sistemática en un servidor de extensa capacidad.

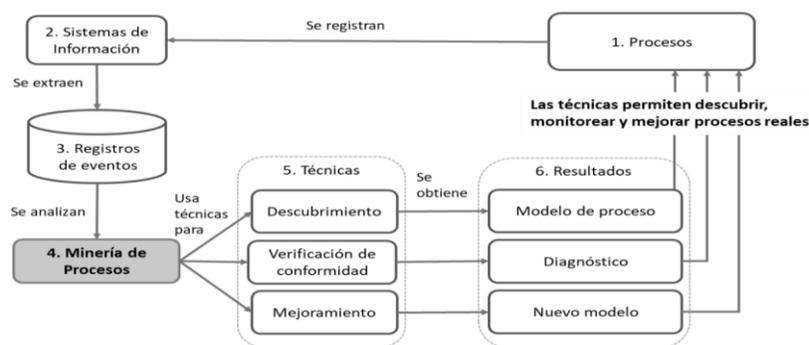
El mayor reto para lograr involucrar dichos datos con la mejora continua de la organización consiste en lograr, de una manera eficaz y eficiente, un alto grado de manipulación e interpretación de los datos almacenados con el fin de obtener respuestas positivas y relevantes. Es por ello que surgieron diferentes metodologías y conceptos basados en esta necesidad, como lo es la minería de procesos.

## Interpretación de datos basada en procesos

Según Agudelo Tobón y Escobar (2007) un proceso se puede definir como “un conjunto de actividades secuenciales o paralelas que ejecuta un productor, sobre un insumo, le agrega valor a éste y suministra un producto o servicio para un cliente externo o interno”. Actualmente, el manejo adecuado y avanzado de procesos con un objetivo de mejora continua está siendo el motor de las organizaciones para el progreso en las industrias de más alto renombre. Dicha mejora se basa en un control adecuado de las actividades realizadas optimizando diferentes variables decisivas con el fin de maximizar en un amplio rango la productividad de dichos procesos. Para ello se hace el uso de análisis de tiempos, rutas, patrones, variantes, entre otros, en donde se pretende recolectar información con el fin de interpretarla, analizarla y, dado el caso, usarla para modificar o replantear procesos.

## Minería de Procesos

La minería de procesos es una disciplina que tiene como objetivo descubrir, monitorear y mejorar procesos a través de la extracción de conocimiento del registro de eventos almacenados en los sistemas de información. (Aguirre y Rincón, 2015). Según textos como el libro Process Mining de Will Van der Aalst, el Manifiesto de minería de procesos de IEEE Task Force on Process Mining, entre otros, se puede definir la minería de procesos como aquel método de análisis de datos generados según eventos de un proceso. A su vez se considera una disciplina que tiene como objetivo recurrir a dichos datos generados con el fin de obtener, monitorear y mejorar procesos, ya sea de un sistema, negocio u otro, a través del análisis exhaustivo del registro de los eventos del proceso, los cuales se almacenan en los sistemas de información especificados por la organización, empresa, etc. La Imagen 1 muestra un resumen sobre cómo se realizan estos análisis.



**Imagen 1.** Realización de la minería de procesos. Elaborado por Van der Aalst (2011).

Para la generación de una correcta minería de procesos, como se muestra en la imagen anterior, se debe iniciar con la generación o definición del proceso, en donde a su vez se va generando la recopilación de los datos que se vayan generando en su desarrollo dentro de un sistema de información. A este tipo de datos, se le conoce como información histórica. De esta información se extraen los eventos que al ser analizados con minería de procesos, sea cual sea la técnica elegida, se genera el modelo real del proceso para luego ser analizado y obtener soluciones o mejoras del proceso. Uno de los objetivos primordiales de la minería de procesos es la realización de un análisis exhaustivo del proceso enfocado en el análisis de datos reales generados según los eventos y el análisis de datos de la red social de un proceso.

De forma general, el primer uso de la minería de procesos es el análisis de datos reales generados según los eventos de un proceso donde en principio las organizaciones poseen un proceso teórico de ejecución, pero a la hora de adquirir los datos, el resultado suele diferir del teórico, y es aquí donde entra a aplicarse los métodos de descubrimiento de la minería de procesos con el fin de hallar el proceso real. Como segundo uso, se encuentra la implementación del método en la auditoría del proceso, realizando la verificación del cumplimiento de las políticas y reglas de negocio. Continuando, se tiene como último aspecto la verificación del desempeño del proceso, en donde se hace uso de conocimiento previo ingenieril para la obtención y verificación de datos como lo son los cuellos de botella, los tiempos reales, tiempo de actividades, entre otros.

Una vez se haya implementado uno o varios de estos usos, los resultados tienden a ser analizados con el fin de obtener alternativas de mejora.

Aplicaciones adicionales de la minería de procesos incluyen el análisis de datos de la red social. Es aquí donde se analizan las personas y los equipos de trabajo que conforman la organización. A cada uno de estos se les desarrolla un análisis similar al realizado al proceso total, ubicando así el sujeto o elemento considerado el cuello de botella y puntos o transiciones en el procesos con demoras. Este tipo de información se toma frecuentemente como base para el desarrollo de propuestas y planes de mejora.

En general, la minería de procesos une el análisis de procesos basado en modelos tradicionales (por ejemplo la simulación) y las técnicas de análisis centradas en los datos, como el aprendizaje automático y la minería de datos. Los datos de eventos combinados con la minería de procesos, permite diagnosticar problemas basados en hechos, no en supuestos. La primera parte de este proyecto, consiste en la aplicación de la minería de procesos a un sistema de manufactura flexible simulado.

### **Casos de implementación de Minería de procesos**

En los últimos años, diversas técnicas de la minería de procesos han sido aplicadas en varias industrias y organizaciones alrededor del mundo, puesto que mediante estas es posible determinar fallas u oportunidades de mejora en procesos complejos que involucran gran cantidad de actividades, personas e información. . Algunos de estos ejemplos son descritos a continuación:

En el año 2005 se realizó un estudio con el departamento de obras públicas de Holanda (Van der Aalst, et al 2005) cuyo objetivo era verificar la aplicabilidad de técnicas de minería de procesos en contextos reales y de un nuevo software para este tipo de problemas llamado ProM. Para ello se tomó como información de entrada aquella relacionada con el tratamiento de facturas en 14279 casos con un total de 147579 eventos o actividades hechas por 487 empleados. El proyecto, se desarrolló con base en tres etapas: la perspectiva del proceso, la perspectiva de la organización y la perspectiva del caso en particular .(Van Der Aalst, 2007).

En la perspectiva del proceso, se ajustó un modelo por medio de una red de petri aplicando el “Algoritmo de minería” (en inglés “Heuristic mining” o “ Mining Algorithm”), un algoritmo de descubrimiento con dos fases donde en la primera se descubren dependencia de actividades por medio de un gráfico y en la segunda relaciones causales a través de un gráfico c-net, para finalmente obtener el modelo deseado (Van Der Aalst, 2016). En cuanto a la perspectiva de la organización, se creó un diagrama social, con lo que se detectó baja participación por parte de altos cargos financieros y demoras en actividades que eran responsabilidad de líderes de proyecto, así como ineficiencias que generaban en ciertas ocasiones que una misma actividad fuese realizada dos veces. Finalmente, en el análisis de la perspectiva del caso, se detectó que aquellas facturas con mayor retraso en su pago eran las que tenían un mayor valor, causando que sus intereses por mora fuesen los más altos. (Van Der Aalst, 2007).

Otra evidencia de la aplicación de la minería de procesos, se presenta en Verenich et al. (2017) donde se empezó el proyecto “*Nirdizati*” destinado a desarrollar una herramienta basada en web para el monitoreo predictivo de procesos de negocios. (Verenich et al, 2018). Posteriormente, se integró la herramienta *Nirdizati* en la plataforma de análisis de procesos basada en la web denominada *Apromore*. A través de esta integración, los usuarios de ésta web pueden usar los registros de eventos almacenados en el repositorio (una carpeta pública de la plataforma) para entrenar una gama de modelos predictivos, y posteriormente usarlos para visualizar varios indicadores de rendimiento de casos de procesos en ejecución de un flujo de eventos. Las predicciones pueden presentarse visualmente en un tablero o exportarse para informes periódicos. Con base en estas predicciones, los gerentes de operaciones pueden identificar posibles problemas desde el principio y tomar medidas correctivas de manera oportuna.

Un documento adicional en el cual se pueden encontrar aplicaciones de la minería de procesos, es el escrito por Jiménez, Zambrano, Aguirre y Trentesaux (2017); “Using process - mining for understating the emergence of self - organizing manufacturing systems .” En este documento se habla acerca de una simulación de un proceso de manufactura flexible basado en las instalaciones de un laboratorio. Según el texto, el experimento realizado se basó en identificar dos tipos diferentes de condiciones, normales y

anormales, donde lo anormal está asociado a una perturbación que ocurre en una de las máquinas del sistema. Es importante mencionar que, los registros de cada evento se almacenaron en un archivo CSV. Al analizar la información obtenida en el transcurso del trabajo, se implementaron distintos algoritmos de análisis de dicha información para la verificación superficial de un sistema, según minería de procesos, para la toma de decisiones en cada uno de los casos atípicos que puedan surgir en el sistema simulado, dando una base para su implementación en contextos reales. Este texto es un referente importante y de mayor impacto en la realización del presente proyecto dado que se tomó como base varios de los puntos, metodologías y análisis propuestos con el fin de darle continuidad y ampliación (desde la perspectiva predictiva) a lo trabajado por estos autores.

Por otro lado, Trentesaux et al. (2013) realizaron un trabajo sobre la Evaluación comparativa de los sistemas flexibles de programación y control de dichos sistemas. Se comparó la producción de diferentes sistemas para un conjunto dado de datos de entrada para mejorar el rendimiento del sistema (Trentesaux et al, 2013). Ellos propusieron un sistema de referencia basado en una célula de producción real ya que, la comunidad de control de producción también ha propuesto puntos de referencia con la intención de permitir la comparación coherente de arquitecturas y sistemas de control de producción. Teniendo en cuenta la dinámica del entorno, esto permite la evaluación de los resultados utilizando las herramientas tradicionales de investigación operativa y la evaluación de la solidez del sistema de control frente a eventos inesperados.

Otro proyecto de investigación relacionado fué el realizado por Verenich, I. (2016). Este estudio, tuvo como objetivo desarrollar un marco general para el monitoreo predictivo de procesos que aborde una variedad de tareas de monitoreo del proceso como lo son predecir el resultado de actividades individuales o del proceso completo, o predecir la ruta de finalización de un evento. Ellos hacen énfasis en que, se han propuesto varias técnicas que son aplicables para conjuntos de datos específicos, pero, la mayoría de los objetivos de predicción se pueden ver como ejemplos específicos de un problema más genérico, es decir, la predicción de una ruta de finalización de un caso determinado y sus características. Su objetivo fue desarrollar un marco general de monitoreo predictivo que pueda ser instanciado para una amplia gama de objetivos de predicción y para diferentes dominios, por ejemplo: Dominios de transporte, seguros y sanidad.

Finalmente, Senderovich, A., Francescomarino, C. D., & Maggi, F. M. (2019) por su parte, realizaron un trabajo en el cual, basan sus investigaciones en el hecho de que las técnicas actuales del monitoreo predictivo de procesos, sólo consideran la información interna que proviene del caso cuyas medidas de interés se desean predecir pero, en muchas ocasiones el resultado de un caso depende la interacción de múltiples casos que se ejecutan simultáneamente. Es decir, es apropiado considerar cada caso de manera conjunta y no unidimensional, por tal motivo, los autores presentan un marco general para la codificación de características que se basa en una representación de espacio bidimensional. Los autores demostraron la utilidad de este marco realizando evaluaciones empíricas con dos conjuntos de bases de datos reales, un proceso hospitalario ambulatorio y un proceso de fabricación respectivamente. Los resultados obtenidos mostraron mejoras del error cuadrático medio al utilizar el marco bidimensional ya que, se predice el tiempo de finalización con respecto a una línea de base donde solo se consideran las características dentro del caso.

### **3. Objetivos**

#### **Objetivo General:**

*Desarrollar una metodología de monitoreo predictivo, haciendo uso de la minería de procesos, para la mejora de la eficiencia de un sistema autoorganizado de manufactura.*

#### **Objetivos Específicos:**

- Caracterizar el funcionamiento del sistema de control de manufactura auto-organizado.
- Diseñar la metodología de monitoreo predictivo para el sistema de control de manufactura auto-organizado, a través de la minería de procesos.
- Aplicar la metodología predictiva propuesta en el simulador del sistema de control de manufactura de la Universidad de Valenciennes, Francia.

- Evaluar el impacto de la metodología en los indicadores de eficiencia del sistema de manufactura.

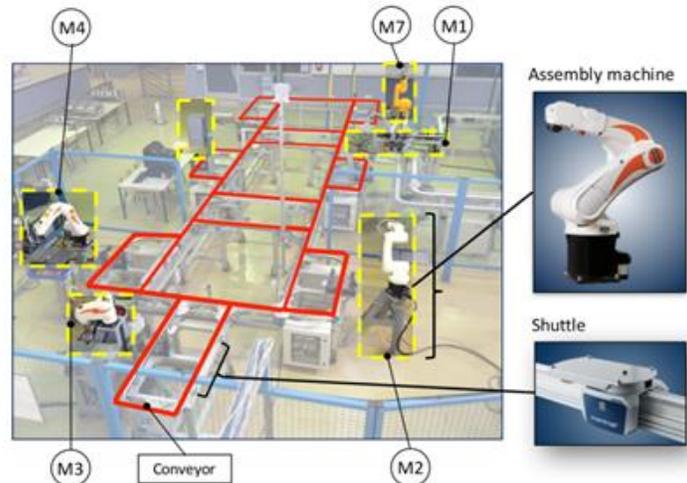
#### 4. Simulador del sistema de manufactura

El caso a trabajar en este documento se centra en el uso del modelo efectuado en el proyecto de Trentesaux et al. (2013), en donde se posee un sistema de manufactura flexible real localizado en la Université de Valenciennes et du Hainaut-Cambrésis (Francia). El sistema tiene como característica esencial ser un sistema de manufactura auto-organizado, lo que significa que posee la habilidad para adaptar dinámicamente su organización y comportamiento, como por ejemplo, crear y eliminar interacciones en determinado proceso, cambiando sus condiciones sin ninguna intervención externa manteniendo su coherencia interna en el ambiente organizacional (Silva & Pierreval, 2013). Es un sistema social, que continuamente genera su comportamiento de acuerdo a las interacciones de los procesos en marcha conformado principalmente por agentes autónomos. (Accard, 2018) .

Un sistema auto-organizado en un contexto de manufactura flexible, como el propuesto en el proyecto de Trentesaux et al. (2013), es aquel que está en capacidad de determinar la secuencia de operaciones a realizar en un caso, el agente que las realiza, y la ruta de cada producto hasta su terminación, generadas a partir de decisiones propias bajo diversos enfoques, como por ejemplo, “primera máquina disponible” (FAM - First Available Machine) o “campos potenciales” (CEP- Potential fields). El sistema a utilizar en este trabajo está basado en la celda de manufactura AIP - PRIMECA la cual es un sistema de ensamblaje basado en componentes industriales como sensores, actuadores, robots, sistemas de manejo de materiales, entre otros componentes, todo ello con el fin de crear un conjunto de procesos de trabajos definidos con una visión de estructuramiento de un sistema de manufactura flexible auto-organizado.

La estructura física general de dicho simulador, la cual se puede observar en la figura 1, se centró en la conformación de un modelo de estaciones de trabajo en donde los principales componentes a tener en cuenta fueron las máquinas, los sensores y los productos. En cuanto a la maquinaria, se tienen siete (7) máquinas de procesamiento (M1, M2, M3, M4, M6 y M7) unidas por bandas transportadoras cuyo fin es realizar cada una de las operaciones a ejecutar. Su tipología está descrita a continuación:

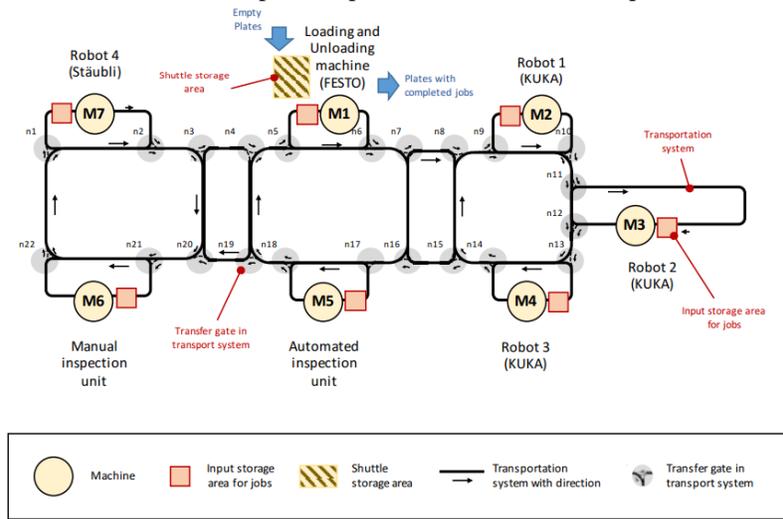
- Máquina de carga /descarga (M1)
- Robot de ensamblaje - KUKA (M2, M3 y M4)
- Unidad manual de inspección (M6)
- Robot de ensamble- STÄUBLI (M7)



**Figura 1.** AIP-PRIMECA: Sistema de manufactura flexible localizado en la Universidad de Valenciennes. (Jimenez, 2017)

Un componente adicional en el sistema de manufactura, son los nodos. Dicha estructura posee una distribución de 29 nodos ( siete máquinas y veintidós puntos de transferencia) los cuales llevan cada uno de los tipos de productos a través de las estaciones. Dentro de estos nodos existen dos clasificaciones generales: aquellos que unen dos rutas diferentes en una (convergentes) y aquellos que dan dos alternativas de escogencia para la ruta (divergentes). Cabe resaltar que dichos nodos se encuentran ubicados a lo largo del

Layout del modelo y no tuvieron modificaciones de ubicación de movimiento, simplemente cumplen la función de marcar el camino a recorrer y recolectar los datos de lo realizado hasta ese punto, como se muestra en la Figura 2. En la figura 3, se muestran los nodos que se utilizaron en el presente proyecto para la generación de registros usados como base para la aplicación de la minería de procesos.



**Figura 2.** Layout celda de manufactura flexible AIP-PRIMECA, Universidad de Valenciennes.(Jimenez, 2017)

El tercer componente de la estructura física del sistema está dado por los productos o trabajos a realizar. El sistema posee la capacidad de realizar siete (7) tipos de productos: B, E, L, T, A, I y P en donde cada uno de estos posee una secuencia de ensamblaje diferente y uso de componentes diferentes según sea el caso. La secuencia de operaciones para cada uno de estos tipos de producto, junto con las máquinas habilitadas para ejecutar dichas operaciones, son presentadas en la figura 4. La definición de estos tipos de producto dada por la secuencia de operaciones, es clave para el análisis del comportamiento del sistema tratado más adelante.

El sistema de la celda de manufactura funciona por medio de un software de controlador que genera la elaboración de los productos requeridos. La ejecución de cada una de las acciones están predeterminadas por la estructura en código realizado en el Software de simulación *NetLogo* (Wilensky, 1999). Dicho software se define como un entorno programable de modelado para simular fenómenos naturales y sociales, además de poseer un manejo adecuado para modelar sistemas complejos que se desarrollan en el tiempo. Los modeladores, contenidos dentro del programa, pueden dar instrucciones a cientos o miles de agentes independientes, entiéndase agentes como los productos a trabajar. La principal ventaja del uso de este software se especifica en la posibilidad de observar y analizar la relación entre el nivel micro del comportamiento de los individuos y los patrones a nivel macro que emergen de la interacción de muchos individuos.

Los datos de entrada usados de la simulación del sistema en *NetLogo* que se usó en el presente trabajo, se basa en cuatro (4) elementos: características del agente, layout, Productos y selección de máquinas (ruta).

Como primer elemento que compone el simulador de *Netlogo*, están las características generales del agente. En dicho archivo existe la posibilidad de realizar el cambio del máximo de elementos posibles dentro del sistema para su evaluación y elaboración, además de las especificaciones del tiempo de entrada entre cada agente y la velocidad de los productos dentro del sistema.

Ahora bien, como segundo elemento, está el layout. Dicho layout contiene la misma estructura que se mencionó anteriormente, manteniendo la base del sistema de manufactura físico con el fin de realizar la simulación lo más certera posible. A su vez, dentro de este se pueden denotar los elementos referentes a cada uno de los componentes de la maquinaria (en amarillo), los nodos y la zonas de inspección, lo cual muestra de manera gráfica y sistemática el funcionamiento del sistema auto-organizado.



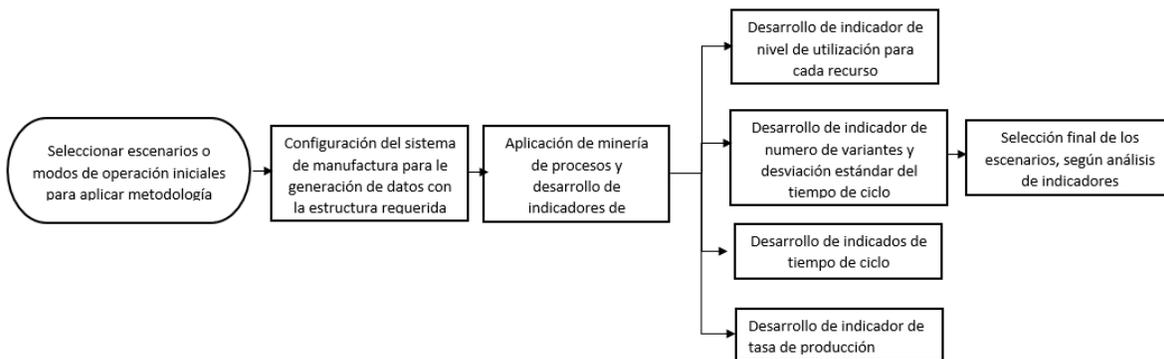
productos se encuentran en línea de espera, y teniendo en cuenta un campo donde todos estos factores se encuentren reunidos y haya menos productos en espera, allí debe ser dirigido el producto.

#### **Escenario 4: Control Alloc - Cyrille Pach**

Este escenario busca tener en cuenta seleccionar la máquina a la cual se va a dirigir el producto que se encuentre dentro de un campo específico potencial delimitado por la distancia y si se encuentra o no ocupado con el fin de que el producto se dirija a la máquina que se encuentre más cerca y que tenga mayor probabilidad de estar disponible en la menor cantidad de tiempo posible, siempre y cuando el atributo del producto pueda ser procesado por esta máquina el producto podrá tomar esa ruta.

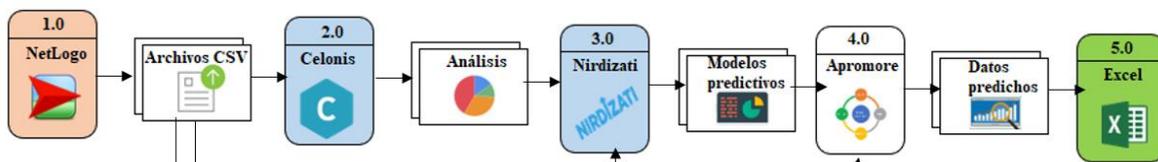
### **5. Diseño de la metodología**

Los pasos relacionados para la aplicación de la metodología de monitoreo predictivo de procesos se muestran y desarrollan a lo largo del presente documento. Se presenta de esta forma para facilidad del lector; sin embargo es preciso resaltar que los pasos propuestos para su aplicación se plasman en la figura 5, figura 12, figura 13, y figura 14; en cada una de dichas figuras se encuentra el orden secuencial la metodología aplicada. A continuación, se muestra la figura 5, que corresponde a la presentación de los primeros pasos de la metodología, los cuales son aplicados en la sección 6.



**Figura 5.** Descripción de la aplicación de minería de procesos en el sistema de manufactura flexible. Elaboración propia

Para la aplicación de la metodología, se utilizaron diversos tipos de software. Como se ha mencionado antes, para la simulación y generación de datos del sistema de manufactura se usó *Netlogo*, para la aplicación de la minería de procesos descrita en la sección 6 se utilizó *Celonis*, para la generación de modelos predictivos de la sección 7 se utilizó la plataforma *Apromore* (junto con la integración de *Nirdizati*) y finalmente para la presentación de resultados en una interfaz determinada se utilizó excel. Cada una de las herramientas mencionadas, requiere la aplicación de habilidades de programación, análisis de datos y manipulación de la información. En este punto, es conveniente aclarar que la aplicación de la metodología dadas las limitaciones de la herramienta utilizada (*apromore*) se da en el sentido de predecir actividades siguientes (next activity) a partir de un registro de eventos. En la figura 6, se muestra el la secuencia de la metodología propuesta a partir de las herramientas utilizadas:



**Figura 6.** Secuencia de la metodología a partir de las herramienta utilizadas. Elaboración propia

Los requerimientos de desempeño de la metodología implican la obtención de modelos de predicción apropiados para al menos uno de los 4 escenarios estudiados. La precisión del modelo para cada uno de los escenarios se midió de acuerdo con un indicador de ajuste del modelo denominado accuracy suministrado por

defecto con la plataforma *apromore* y con tablas comparativas entre actividades predichas y actividades ejecutadas. A modo de prueba de rendimiento de la metodología se validó cada uno de los modelos con datos que no fueron usados para su entrenamiento, obteniéndose los resultados descritos en la sección 7.2. Finalmente, en cuanto a restricciones de la metodología, se halló relación entre la imposibilidad de generar un modelo con predicciones apropiadas y la varianza del tiempo de ciclo de los casos, entre mayor variabilidad en el tiempo de ciclo de los productos, existe mayor dificultad de encontrar patrones de comportamiento y de entrenar modelos de predicción.

Cada uno de los análisis requeridos para el desarrollo de la metodología, en el caso particular del sistema de manufactura, se irán desarrollando paulatinamente a medida que esta se aplicada.

## 6. Análisis descriptivo con minería de procesos

Para cada uno de los escenarios descritos en la sección anterior, se ejecutó una misma orden de producción donde se generaron registros de las diferentes operaciones del sistema de manufactura. Se simuló la producción de 1400 unidades formadas de 200 elementos de cada uno de los 7 tipos de producto. Cada uno de los registros de eventos de cada escenario fue analizado por medio de la minería de procesos haciendo uso del software de análisis *Celonis*, con el que se obtuvieron Dashboards, histogramas y gráficos tipo Process mining. A modo ilustrativo, en la figura 7 se muestra el diseño del dashboard para el escenario cyrille. Los dashboards de los demás escenarios se muestran en el documento de Anexos, ANEXO 1.

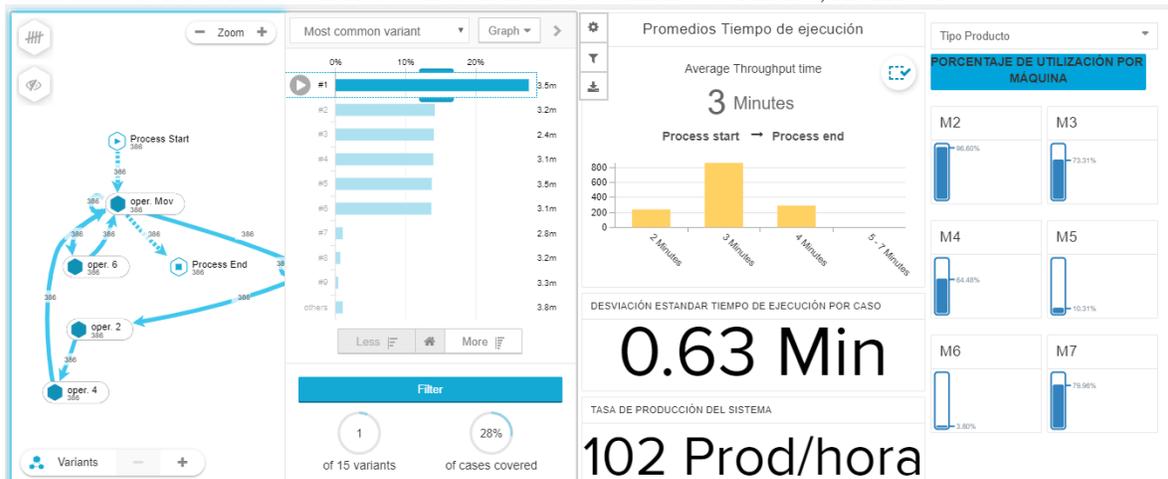


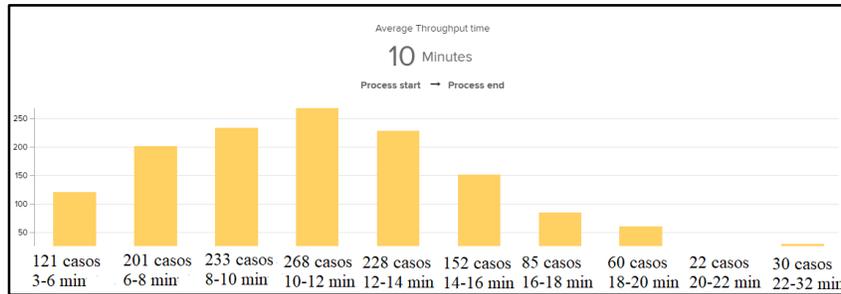
Figura 7. Dashboard descriptivo por minería de procesos para el escenario cyrille. Elaboración propia.

Este tipo de ilustraciones, permitió realizar una comparación del comportamiento de los escenarios con base en diferentes aspectos e indicadores que se muestran detalladamente en la tabla 5. Sin embargo, en principio se realizó un análisis general del comportamiento del sistema simulado para cada uno de los escenarios, el cuál es mostrado a continuación:

### 6.1 Elaboración de análisis por escenario.

#### Escenario 1 : Randomly

En este escenario, los 1400 productos que componen la orden de producción implementada fueron ejecutados en un tiempo de 45 horas. El tiempo de ciclo promedio de cada uno de los casos es de 10 minutos, con variaciones entre los 3 minutos hasta los 32 minutos en algunos casos, siendo el escenario con mayor variabilidad como se ve en la tabla 5. Estos tiempos de ejecución involucran un promedio de 227 eventos por caso (producto) entre operaciones en la máquinas y movimientos de traslado dentro del sistema. Un histograma representativo relacionado con la duración de los casos, se muestra en el dashboard del escenario, en el documento de Anexos, ANEXO 1.1. Sin embargo a modo ilustrativo, a continuación en la figura 8 se presentan rangos de tiempo incluidos en el histograma con la respectiva proporción de casos implicados.



**Figura 8.** Histograma de tiempos de procesamiento para el escenario Randomly. Elaboración propia

Los casos con mayor tiempo de ciclo corresponden a 30 productos en su mayoría de tipo B, A y E a los que se les detectó retrasos en el tiempo de ejecución en el traslado entre recursos antes de ejecutar las operaciones 1 y 5. En estos casos, cada una de las actividades son ejecutadas en máquinas diferentes o al menos existe un traslado entre actividades que aumenta su tiempo de procesamiento. Por otra parte, los casos con baja duración de tiempo de procesamiento corresponden principalmente a productos tipo P, T e I que aunque realizan sus operaciones en máquinas diferentes (es decir se generan más traslados), dichos traslados son rápidos y no aumentan en gran medida el tiempo del producto en el sistema.

En el escenario, se encontraron 122 diferentes variantes, siendo un resultado razonable, dadas las características de aleatoriedad. Debido a la cantidad de variantes, ninguna de ellas involucra un porcentaje considerable de casos asociados, por lo que en el escenario particular no se hace conveniente el análisis de alguna variante en particular en cuanto a mayor o menor tiempo de procesamiento. Sin embargo, mediante un análisis conjunto de las variantes que aumentan el tiempo de procesamiento, se encontró que los retrasos se dan principalmente por los tiempos de traslado existentes hacia las operaciones 1, 2 y 5. Por otro lado, un análisis conjunto de las variantes con menor tiempo de procesamiento permitió detectar una particularidad en ellas: existen varias operaciones que se ejecutan en una misma máquina o recurso, disminuyendo los tiempos de traslado.

En cuanto al análisis de los recursos, el porcentaje de utilización cada una de las máquinas en el sistema se presenta en la tabla 5. Sin embargo, como complemento al presente análisis, a continuación se presenta un listado de la cantidad de eventos o actividades ejecutados por cada una:

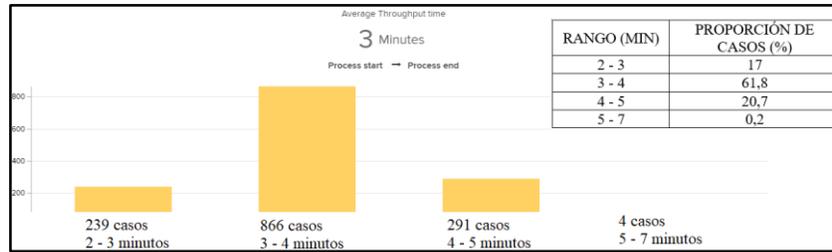
Máquina	Número de eventos asociados	Tiempo promedio de casos que usan la máquina o recurso respectivo
2	2282	11.2 minutos
3	2256	11.2 minutos
4	1671	11.2 minutos
5	810	10.8 minutos
6	590	11.2 minutos
7	1591	11.6 minutos

**Tabla 1.** Eventos asociados por recurso - escenario Randomly. Elaboración propia.

En el ANEXO 1.1. se muestra un dashboard general, resultado del análisis de procesos, que resume información adicional del escenario, y que se analiza en el literal 5.2. Comparación entre escenarios.

### Escenario 2: Control Alloc - Cyrille Pach

En este escenario, los 1400 productos que componen la orden de producción implementada fueron ejecutados en un tiempo de 13,79 horas. El tiempo de ciclo promedio de cada uno de los casos es de 3 minutos, con variaciones entre los 2 minutos hasta los 7 minutos en algunos casos. Estos tiempos de ejecución involucran un promedio de 26 eventos por caso entre operaciones en la máquinas y movimientos de traslado dentro del sistema. Un histograma representativo relacionado con la duración de los casos, se muestra en el dashboard del escenario, en el documento de Anexos, ANEXO 1.2. Sin embargo a modo ilustrativo, a continuación en la figura 9 se presentan rangos de tiempo incluidos en el histograma con la respectiva proporción de casos implicados.



**Figura 9.** Histograma del tiempo de procesamiento para el escenario Control Alloc - Cyrille Pach. Elaboración propia

Los casos con mayor tiempo de ciclo corresponden a 4 productos de tipo A y E a los que se les detectó retrasos en el tiempo de ejecución al momento de ingresar al sistema y al momento de encontrar una máquina disponible para la ejecución de su primera actividad (Oper. 1). Por esta razón, al momento de su ingreso al sistema permanecieron en movimiento (trasladándose) más de lo esperado. Por otra parte, los casos con baja duración de tiempo de procesamiento corresponden principalmente a productos tipo I, los cuales poseen tiempos bajos en movimiento a comparación de los productos A y E.

En el escenario, se encontraron 15 diferentes variantes. Algunas de ellas poseen pocos productos involucrados (menos del 1% de los casos) como en el caso de los 4 casos de productos A y E que corresponden a los casos con mayor duración. La variante a la cual sus productos asociados poseen un mayor tiempo de procesamiento (aproximadamente 3.5 minutos) y que involucran un porcentaje considerable de casos (28% para el caso particular) es:

Oper. 1 → Oper. 2 → Oper. 4 → Mov. → Oper. 6

Dicha secuencia de operaciones corresponde a tipos producto T y P. Otras variantes con alto tiempo de procesamiento involucran productos tipo A. El retraso de la secuencia, está dado por el tiempo de traslado al momento de la entrada al sistema y el comienzo de la ejecución de la operación 1. Por otro lado, la variante con menor tiempo de ciclo en sus productos (asociada a los productos I, P y T) corresponde a:

Oper. 1 → Oper. 5 → Mov. → Oper. 3 → Mov. → Oper. 6

Nótese que los resultados de las variantes, coinciden con los tipos de producto con mayor demora registrada en el histograma de la figura 9.

En cuanto al análisis de los recursos, el porcentaje de utilización cada una de las máquinas en el sistema se presenta en la tabla 5. Sin embargo, como complemento al presente análisis, a continuación se presenta un listado de la cantidad de eventos o actividades ejecutados por cada una:

Máquina	Número de casos	Tiempo promedio de casos que usan la máquina o recurso respectivo
2	2397	3.4 minutos
3	1819	2.8 minutos
4	1600	3.1 minutos
5	1023	3.1 minutos
6	377	2.8 minutos
7	1984	2.8 minutos

**Tabla 2.** Eventos asociados por recurso - escenario Control Alloc - Cyrille Pach. Elaboración propia.

En el ANEXO 1.2. se muestra un dashboard general, resultado del análisis de procesos, que resume información adicional del escenario, y que se analiza en el literal 5.2. Comparación entre escenarios.

### Escenario 3: First Available Machine (FAM)

En este escenario, la ejecución de la orden de producción fue de 26 horas. El tiempo de ciclo promedio de cada uno de los casos es de 6 minutos, con variaciones entre los 2 minutos hasta los 14 minutos en algunos casos. Estos tiempos de ejecución involucran un promedio de 88 eventos por caso entre operaciones en la máquinas y movimientos de traslado dentro del sistema. Un histograma representativo relacionado con el tiempo de ciclo de los casos, se muestra en el dashboard del escenario, en el documento de Anexos, ANEXO 1.3. Sin embargo a modo ilustrativo, a continuación en la figura 10 se presentan rangos de tiempo incluidos en el histograma con la respectiva proporción de casos implicados.

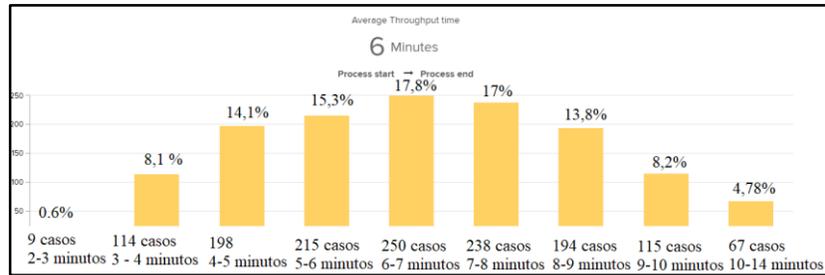


Figura 10. Histograma del tiempos de procesamiento para el escenario FAM. Elaboración propia.

Los casos con mayor tiempo de ciclo corresponden a 4 productos de tipo A, B y L, los cuales realizan cada una de sus operaciones en diferentes máquinas. Esto genera mayor cantidad de traslados, que aumentan el tiempo de los productos en el sistema, especialmente aquellos generados antes de realizar las operaciones 1 y 2. Por otra parte, los casos con bajo tiempo de procesamiento corresponden principalmente a productos tipo I, P y T, que aunque también cada una de sus actividades en una máquina diferente, el tiempo de traslado entre máquinas es significativamente más reducido.

En el escenario, se encontraron 6 diferentes variantes. Cada una de las variantes está asociada a un tipo de producto excepto una variante que es compartida por dos de ellos (producto P y T) los cuales comparten las misma operaciones. La variante a la cual sus productos asociados poseen un mayor tiempo de procesamiento (aproximadamente 8.4 minutos) y que involucran un porcentaje considerable de casos (14% para el caso particular), corresponde al tipo de producto B, se presenta a continuación:

Oper. 1 → Mov. → Oper. 2 → Mov. → Oper. 3 → Mov. → Oper. 5 → Mov. → Oper. 6

Por otro lado, la variante con menor tiempo de ciclo en sus productos (asociada a los productos P y T) corresponde a

Oper. 1 → Mov. → Oper. 2 → Mov. → Oper. 4 → Mov. → Oper. 6

Dado que cada variante tiene uno o máximo dos productos asociados al recorrido como se mencionó antes, nótese que dichas variantes y productos asociados coinciden con los resultados de casos (productos) con mayor duración y menor duración (o tiempo de ciclo) encontrados a partir del histograma. Finalmente continuando con la estructura del análisis general que se ha venido desarrollando para cada uno de los productos, a continuación de muestras los recursos con su respectiva cantidad de casos atendidos:

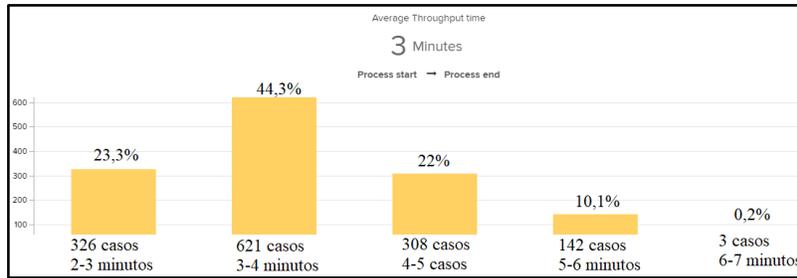
Máquina	Número de eventos asociados	Tiempo promedio de casos que usan la máquina o recurso respectivo
2	2998	6.2 minutos
3	2664	6.3 minutos
4	1920	6.4 minutos
5	1400	6.2 minutos
6	0	--
7	218	6.6 minutos

Tabla 3. Eventos asociados por recurso - escenario FAM. Elaboración propia.

En el ANEXO 1.3. se muestra un dashboard general, resultado del análisis de procesos, que resume información adicional del escenario, y que se analiza en el literal 5.2 (*Comparación entre escenarios*).

#### Escenario 4: Potential fields (CEP)

En este escenario, la ejecución de la orden de producción fue de 14.4 horas. El tiempo de ciclo promedio de cada uno de los casos es de 3 minutos, con variaciones entre los 2 minutos hasta los 7 minutos en algunos casos. Estos tiempos de ejecución involucran un promedio de 28 eventos por caso entre operaciones en la máquinas y movimientos de traslado dentro del sistema. Un histograma representativo relacionado con el tiempo de ciclo de los casos, se muestra en el dashboard del escenario, en el documento de Anexos, ANEXO 1.4. Sin embargo a modo ilustrativo, a continuación en la figura 11 se presentan rangos de tiempo incluidos en el histograma con la respectiva proporción de casos implicados.



**Figura 11.** Histograma del tiempos de procesamiento para el escenario CEP. Elaboración propia.

Los casos con mayor tiempo de ciclo corresponden a 3 productos de tipo A y L (2 productos tipo A y un producto tipo L), los cuales aunque realizan varias de sus operaciones en una misma máquina (lo cual reduce sus tiempos de traslado), poseen retrasos al momento de ingresar al sistema y ejecutar la operación 1. Esto implica que dichos productos permanecen en movimiento por aproximadamente 2 minutos (causantes del retraso de tiempo) antes de ejecutar la primera operación. Por otra parte, los casos ejecutados en menos tiempo corresponden principalmente a productos tipo P e I, que a pesar de tener bajos tiempos de procesamiento, encuentran de igual forma retrasos al ingreso al sistema y en el tiempo de traslado hacia la ejecución de la primera operación.

En el escenario, se encontraron 14 diferentes variantes. Aquellas con mayor tiempo de procesamiento poseen pocos productos involucrados (menos del 1% de los casos) como en el caso de los 3 casos de productos A y L mencionado anteriormente. La variante a la cual sus productos asociados poseen un mayor tiempo de procesamiento (aproximadamente 5 minutos) y que involucran un mayor porcentaje (7% para el caso particular) se asocia exclusivamente al tipo de producto A. Dicha variante se presenta a continuación:

**Oper. 1 → Oper. 2 → Mov. → Oper. 4 → Mov. → Oper. 5 → Oper. 3 → Mov. → Oper. 6**

En la anterior variante, el retraso de tiempo se encontró en el traslado entre máquinas hacia la operación 4. Por otro lado, la variante con menor tiempo de ciclo en sus productos, que involucra a un porcentaje aceptable de casos (14% de los casos específicamente) está asociada a productos de tipo P y T (198 productos tipo P y 2 tipo T) y tiene un tiempo de ciclo promedio de 2.3 minutos. Dicha variante se presenta a continuación:

**Oper. 1 → Oper. 2 → Mov. → Oper. 4 → Mov. → Oper. 6**

Realizando la comparación entre los productos asociados a las variantes anteriores, es posible encontrar coincidencias respecto a los productos con mayor y menor duración encontrados a través de análisis del histograma de la figura 11 mostrado anteriormente. Finalmente continuando con la estructura del análisis general que se ha venido desarrollando para cada uno de los productos, a continuación de muestras los recursos con su respectiva cantidad de casos atendidos:

Máquina	Número de eventos asociados	Tiempo promedio de casos que usan la máquina o recurso respectivo
2	2403	3.7 minutos
3	2095	2.8 minutos
4	1697	3.2 minutos
5	1298	3.2 minutos
6	102	2.5 minutos
7	1605	3.7 minutos

**Tabla 4.** Eventos asociados por recurso - escenario CEP. Elaboración propia.

En el ANEXO 1.4. se muestra un dashboard general, resultado del análisis de procesos, que resume información adicional del escenario, y que se analiza en el literal 5.2 (*Comparación entre escenarios*)

## 6.2. Comparación de indicadores entre escenarios:

Escenario	Tasa de producción (prod/min)	Cantidad de variantes	Tiempo de ciclo (segundos)	Desviación estándar del tiempo de ciclo	Porcentaje (%) de utilización por máquina					
					M2	M3	M4	M5	M6	M7
Control Alloc - Cyrille Pach	1.69	15	182	27.72	96,60	73,31	64,48	10,31	3,80	79,95
First Available Machine	0.9	6	370	116.15	64,04	56,90	41,01	7,48	0	4,66
Potencial fields	1.62	14	184	48.95	92,69	80,81	65,46	12,52	0,98	61,91
Randomly	0.52	122	658	258.85	28,38	28,06	20,78	2,52	1,83	19,79

**Tabla 5.** Cuadro comparativo de indicadores entre los diferentes escenarios obtenidos mediante la minería de procesos. Elaboración propia.

### Randomly

Este escenario posee la menor tasa de producción y el mayor tiempo de ciclo promedio por caso. La desviación estándar que presenta corresponde al 39,33% de su media, siendo también la más alta tanto en valor relativo como nominal. De igual forma, el aprovechamiento de los recursos disponibles (máquinas) es bajo en comparación a los demás escenarios, teniendo como el recurso mejor aprovechado la máquina 2 con el 28,38% del tiempo total, valor similar al de otras máquinas como 3, 4 y 7. Dado el método aleatorizado de selección de recursos o máquinas para la ejecución de operaciones, los resultados obtenidos coinciden con lo esperado.

La alta variabilidad del escenario, representada por la desviación estándar y el mayor número de variantes entre todos los escenarios, provee indicios sobre la dificultad de realizar predicciones adecuadas. Adicionalmente el bajo desempeño del escenario medido con los indicadores descritos muestra que la aplicación del escenario, limitaría la eficiencia del sistema por lo que, dadas sus condiciones y características actuales no sería el escenario más apropiado, para la implementación de un dashboard predictivo.

### Control Alloc - Cyrille Pach

En este escenario, es posible observar que la varianza del tiempo de ciclo por caso corresponde al 15,23% de la media. Un valor relativamente pequeño que, permite considerarlo como un candidato adecuado para la aplicación de técnicas predictivas. La tasa de producción y el alto porcentaje de utilización por máquina, llegando al 96% para la máquina 2 con 2397 eventos ejecutados, muestra que es uno de los escenarios más eficientes en comparación con los demás estudiados en el presente proyecto.

### First Available Machine

En el caso de la configuración del escenario First Available Machine, es posible evidenciar que sus casos poseen en promedio el segundo mayor tiempo de ciclo de los 4 escenarios estudiados, con una varianza que representa el 31,39% de su media. La tasa de producción del sistema completo es baja en comparación con los escenarios potencial fields y Control Alloc – Cyrille Pach (aproximadamente un 55% de las tasas de producción de los escenarios mencionados). Dicho aumento en la duración promedio de casos, se da por la baja utilización de M7 con solo 4,66%, a pesar de que dicho recurso puede realizar parte de las operaciones más requeridas en los productos como operación 1 y operación 2.

Aunque por su aceptable variabilidad dada por la desviación estándar y el menor número de variantes de todos los escenarios, puedan desarrollarse modelos predictivos confiables para este escenario, debido a los indicadores mencionados antes, como tasa de producción y el tiempo de ciclo promedio por caso, su aplicación en un contexto real o simulado no generará los mejores resultados (con base en los indicadores establecidos) en comparación con otros escenarios ya descritos.

### Potential fields

Potential fields tiene un comportamiento similar al de Control Alloc – Cyrille Pach excepto en el nivel de variabilidad dado por la desviación estándar del tiempo de ciclo por caso y la cantidad de variantes. En este

caso, su desviación estándar de 48,95 seg, corresponde al 26,6% del tiempo promedio por caso. La cantidad de variantes es más del doble que el escenario FAM y Control Alloc - Cyrille Pach. Dada su alta tasa de producción respecto a otros escenarios y al ser el escenario con segunda menor desviación respecto de la media, es conveniente generar y entrenar modelos de predicción. Se espera que a través del entrenamiento de modelos predictivos sea posible generar datos adecuados y precisos del comportamiento que tendrá el sistema.

### **6.3. Análisis general**

De acuerdo al análisis por escenario y al comparativo de indicadores entre escenarios (secciones desarrolladas anteriormente) es posible caracterizar el comportamiento del sistema a partir de factores comunes observados en todos los escenarios. Por ejemplo, se observó que el elemento que añade tiempo al procesamiento de los productos son los traslados entre máquinas para la ejecución de sus operaciones respectivas; principalmente traslados al ingreso al sistema antes de realizar la operación 1, y el traslado que ocurre antes de realizar la operación 2. Estas situaciones, permitieron caracterizar y entender de una mejor manera la implicaciones que tiene el diseño de los productos, de la orden de producción y de la configuración dada a las máquinas en el simulador. Del mismo modo el descubrimiento de dichos factores de comportamiento en el sistema permitió verificar de primera mano la utilidad de la minería de procesos en un contexto de manufactura, cuyo único estudio al respecto (Jiménez et al., 2018) es la base del presente proyecto.

Profundizando en la situación presentada como ejemplo en el párrafo anterior sobre los elementos que aumentan tiempo al proceso, se encontró que los únicos recursos que están habilitados para la ejecución de la operaciones 1 y 2 son M2, M3 y M7; y que los 3 recursos pueden realizar ambas operaciones. Teniendo en cuenta que todos los tipos de producto (y por consiguiente el 100% de los casos estudiados) requieren al principio la ejecución de la operación 1 y el 71% de los casos requieren posteriormente la operación 2, se puede afirmar que las máquinas M2, M3 y M7 son recursos limitantes. Por ejemplo, en un momento dado pueden haber  $n$  ( $n > 3$ ) productos requiriendo de la operación 1, por lo que  $n-3$  productos estarían en movimiento esperando para usar dichos recursos. Por el contrario, recursos como M5 y M6 son recursos con holguras en su tiempo de uso debido a que ejecutan una única actividad (operación 6) cuya duración es de 5 segundos, en comparación con el resto de actividades que se ejecutan en 20 segundos. Por su parte, el recurso M4, a pesar de que es el único habilitado para realizar la operación 3 (lo que implica que el 57% de los casos deben pasar por el) no es un recurso limitante debido a que la operación 3 se ejecuta a lo sumo 2 veces por producto, mientras que las operaciones 1 y 2 se ejecutan hasta 5 veces por caso.

El nivel de utilización de los recursos presentado en la tabla 5 y en los dashboards del ANEXO 1, en el documento de Anexos, coincide con lo mencionado anteriormente. Mientras que en términos generales, los recursos con mayor porcentaje de utilización son M2, M3, M4, y M7, aquellos que menos son utilizados son M5 y M6. Dicho resultado, puede sugerir en principio, la existencia de una mejora en el desempeño del sistema si se redistribuyen las actividades asignadas a cada recurso, o preferiblemente si se habilitan cada uno de los recursos para la ejecución de todas las operaciones, que sin embargo, en sistemas de manufactura no flexibles implicaría costos de inversión y restricciones adicionales para cada caso particular. Esta redistribución de actividades, generaría menores tiempos de desplazamiento y (dependiendo del escenario utilizado) aumentos en la tasa de producción.

La asociación entre los tipos de producto y su tiempo de procesamiento, es otro factor muy relevante a analizar. En todos los escenarios los tipos de producto con menor tiempo de procesamiento son P, T e I. Esto no sorprende puesto que, son los productos que menor número de operaciones necesitan (5 operaciones). Sin embargo, no solo el número de operaciones de un tipo de producto es determinante para señalarlo como el de menor tiempo de procesamiento; también se debe tener en cuenta el escenario en el que se ejecuta una orden de producción. Por ejemplo, la variante con mayor tiempo de procesamiento en el escenario Control Alloc - Cyrille Pach involucra productos tipo P y T indicando que, para el escenario particular se le está dando prioridad en algunos casos a aquellos productos con mayor número de actividades. Por otra parte, los tipos de productos con mayor tiempo de ejecución en todos los escenarios son A, B, E y L; los cuales son como se esperaba, aquellos con mayor número de actividades asociadas.

Finalmente, y de acuerdo a la tabla 5 es posible afirmar que los mejores escenarios de acuerdo a su tasa de producción y su varianza relativa frente a la media del tiempo de ejecución por caso son Control Alloc - Cyrille Pach y Potential fields. Adicionalmente, son aquellos escenarios que tienen un mejor aprovechamiento de los recursos, visto este a partir del nivel de utilización de las máquinas. Se espera que la aplicación del análisis predictivo en estos escenarios, genere información apropiada y relevante para el desarrollo del proceso. Un escenario adicional, del que se esperaron resultados apropiados es FAM, que aunque posee casi la mitad de la tasa de producción de los escenarios mencionados, puede ser un escenario de estudio que genere conclusiones interesantes. En cuanto al escenario randomly, dada su alta variabilidad, mostrada tanto en cantidad de variantes como en la desviación estándar del tiempo de ciclo de sus casos y la baja tasa de producción, es normal esperar predicciones no confiables. En la sección siguiente, se presenta al análisis predictivo para cada uno de los escenarios.

## 7. Análisis predictivo con minería de procesos

Para desarrollar la metodología predictiva de cada escenario se hizo necesario hacer el análisis de acuerdo a tres muestras de cada uno de los escenarios: la primera con los 28 primeros registros de productos terminados, la segunda con los 49 primeros registros de productos terminados y la tercera con todos los registros de productos terminados. Las primeras dos pruebas se realizaron con números múltiplos de 7 ya que el sistema realiza 7 tipos de productos. La realización de estas muestras tiene como fin evaluar los modelos en los resultados y definir todo un modelo concreto para el escenario como tal. Cabe resaltar que dicho análisis está basado en predecir la siguiente actividad a realizar en la línea de producción del producto en elaboración.

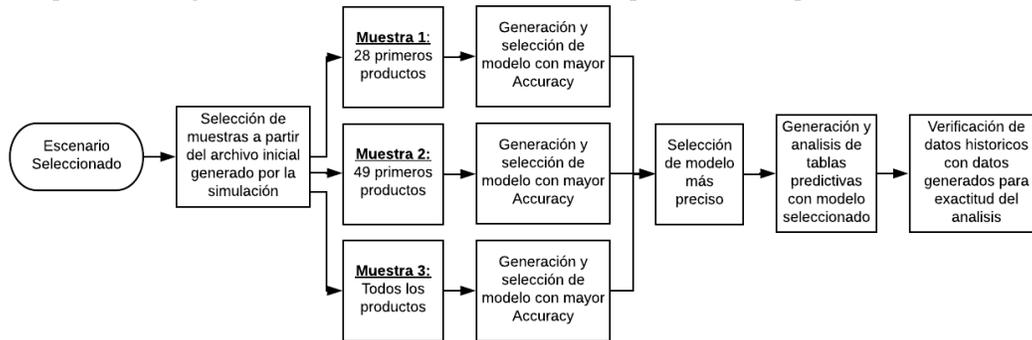


Figura 12. Proceso a realizar para el análisis por escenario. Elaboración propia.

Como se puede observar en la Figura 12, el proceso inicia en la selección de uno de los cuatro escenarios definidos anteriormente. Una vez seleccionado dicho escenario se procede a generar los datos históricos del procesamiento de los 1400 productos. Al obtener los resultados de las rutas de procesamientos para el desarrollo de cada uno de los productos de dicho escenario, se procede a generar las tres muestras requeridas (28, 49 y completo). Cada una de las muestras se analiza según diferentes modelos (Véase en el documento de Anexos, ANEXO 2) y se procede a verificar los resultados, eligiendo así el modelo predictivo con mayor nivel de Accuracy (exactitud). Posteriormente, se realiza una comparación de los 3 modelos (un modelo por cada muestra de 28, 29 y 1400 productos) por escenario, y se escoge (por cada escenario) el más apropiado de acuerdo con los niveles de exactitud arrojados por el análisis. Este modelo seleccionado para cada escenario es usado para generar los datos predictivos. Dichos datos se comparan con los datos producidos originalmente en la simulación (datos históricos) y se analizan para comprobar la veracidad de la predicción realizada por el modelo seleccionado.

### 7.1. Generación y selección de modelo por escenario.

Para escoger el modelo por escenario, como se mencionó anteriormente se usó el criterio de Accuracy (Exactitud) realizando un análisis gráfico y numérico con cada uno. En análisis numérico, se escogió aquel modelo cuyo nivel de Accuracy se acercara más a 1. Con el análisis gráfico se realizó un comparativo con todos los modelos disponibles, escogiendo aquel cuya gráfica fuese la más similar a la de los datos originales,

es decir aquella gráfica cuya pendiente no tuviera significativa variación respecto a la original. Los resultados obtenidos por escenario fueron los siguientes:

**6.1.1.** Para iniciar, se realizó el análisis de manera gráfica con la intención de comparar todos los modelos planteados. A continuación los resultados obtenidos:

<b>ENTRENAMIENTO DE MODELOS POR MUESTRA</b>			
<b>RESULTADOS GRÁFICOS</b>			
	<b>MUESTRA COMPLETA</b>	<b>MUESTRA 28 PRODUCTOS</b>	<b>MUESTRA 49 PRODUCTOS</b>
<b>RANDOMLY</b>	Los modelos escogidos fueron Prefix Length based y State based debido a que hay una gran cercanía entre sus pendientes y adicional, en conjunto, son los modelos que gráficamente están en mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 3.1.)	Los modelo escogidos fueron Clustering y State based debido a que hay una gran cercanía entre sus pendientes y la correspondiente a los datos originales. Además, ambos modelos están, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 3.2.)	Los modelos escogidos fueron Prefix Length based debido a que hay una gran cercanía entre su pendiente y la correspondiente a los datos originales. Además, el modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 3.3.)
<b>FIRST AVAILABLE MACHINE (FAM)</b>	Los modelos escogidos fueron Prefix Length based y State Based dada la cercanía de sus pendientes y adicional, en conjunto, son los modelos que gráficamente están en mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 4.1.)	El modelo escogido fue State Based dada la cercanía de su pendiente a la pendiente de los datos originales. Además, el modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 4.2.)	El modelo escogido fue State Based dada la cercanía de su pendiente a la pendiente de los datos originales. Además, el modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 4.3.)
<b>POTENTIAL FIELDS (CEP)</b>	Los modelos escogidos fueron None y Clustering dada la cercanía de sus pendientes y adicional, en conjunto, son los modelos que gráficamente están en mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 5.1.)	Los modelos escogidos fueron Clustering y None dada la cercanía de sus pendientes a la pendiente de los datos originales. Además, ambos modelos están, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 5.2.)	El modelo escogido fue None dada la cercanía de su pendiente a la pendiente de los datos originales. Además, el modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 5.3.)
<b>CONTROL ALLOC - CYRILLE PACH</b>	El modelo escogido fue Clustering dado que es el modelo que gráficamente está en mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 6.1.)	El modelo escogido fue Clustering debido a que hay una gran cercanía entre su pendiente y la correspondiente a los datos originales. Además, dicho modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 6.2.)	El modelo escogido es Clustering debido a que hay una gran cercanía entre su pendiente y la correspondiente a los datos originales. Además, dicho modelo está, gráficamente, con mayor cercanía al punto más exacto, 100% de acc. (Véase en el documento de Anexos, ANEXO 6.3.)

**Tabla 6.** Entrenamiento de los modelos , en las muestras, para la generación del "Next Activity" por escenario. Análisis Gráfico. Elaboración propia.

6.1.2. Seguido al análisis gráfico, se procedió a realizar un análisis numérico, del cual a continuación se muestran los resultados correspondientes basados en el dato del Accuracy obtenido:

<b>ENTRENAMIENTO DE MODELOS POR MUESTRA</b>			
<b>RESULTADOS <u>NUMÉRICOS</u></b>			
	<b>MUESTRA COMPLETA</b>	<b>MUESTRA 28 PRODUCTOS</b>	<b>MUESTRA 49 PRODUCTOS</b>
<b>RANDOMLY</b>	Los modelos escogidos fueron Prefix Length based y State based dado a que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 7.1.)	Los modelos escogidos fueron Clustering y State based do a que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 7.2.)	Los modelos escogidos fueron Prefix Length based dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 7.3.)
<b>FIRST AVAILABLE MACHINE (FAM)</b>	Los modelos escogidos fueron Prefix Length based y State Based dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 8.1.)	Los modelos escogidos fueron State Based dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 8.2.)	Los modelos escogidos State Based e Index Based dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 8.3.)
<b>POTENTIAL FIELDS (CEP)</b>	Los modelos escogidos fueron none y Last State dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 9.1.)	Los modelos escogidos fueron Clustering y None dada su cercanía en el nivel de Accuracy a 1. Estos resultados concuerdan con los del análisis gráfico. (Véase en el documento de Anexos, ANEXO 9.2.)	El modelo escogido fue Clustering y none debido a que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 9.3.)
<b>CONTROL ALLOC - CYRILLE PACH</b>	El modelo escogido fue Clustering dado que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 10.1.)	El modelo escogido fue Clustering ya que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 10.2.)	El modelo escogido fue Clustering debido a que su nivel de ACC es muy cercano a 1 lo que revela gran semejanza con los datos originales, y además de esto concuerda con el análisis gráfico. (Véase en el documento de Anexos, ANEXO 10.3.)

**Tabla 7.** Entrenamiento de los modelos , en las muestras, para la predicción del “Next Activity” por escenario. Análisis Numérico. Elaboración propia.

Una vez obtenidos los resultados de los análisis numéricos y gráficos, se realizó la selección final del modelo más preciso por escenario. Los modelos seleccionado y el análisis individual se muestran en la siguiente tabla:

ENTRENAMIENTO DE MODELOS POR MUESTRA	
<u>SELECCIÓN FINAL</u>	
<b>RANDOMLY</b>	<b>Prefix Length based</b> (Véase en el documento de Anexos, ANEXO 11.1.)
<b>FIRST AVAILABLE MACHINE (FAM)</b>	<b>State Based</b> (Véase en el documento de Anexos, ANEXO 11.2.)
<b>POTENTIAL FIELDS (CEP)</b>	<b>None</b> (Véase en el documento de Anexos, ANEXO 11.3.)
<b>CONTROL ALLOC - CYRILLE PACH</b>	<b>Clustering</b> (Véase en el documento de Anexos, ANEXO 11.4.)

**Tabla 8.** Selección final de los modelos para la generación del “*Next Activity*” por escenario. Elaboración propia.

Ahora bien, como se pudo observar durante el desarrollo y la selección de los modelos por escenario, se hizo necesaria el orden especificado en la figura 10, en donde la estructura indicaba tomar los datos generados por el proceso de manufactura , seleccionar tres muestras por cada escenario, analizarlas y seleccionar el modelo más preciso. Al realizar el paralelo entre los resultados gráficos y los resultados numéricos se evidenciaron tres aspectos:

1. Nivel de Accuracy diagramado
2. Nivel de Accuracy numérico
3. Semejanza y repetición en la selección del modelo tanto numérico como gráfico.

Dichos aspectos, tuvieron como resultado una similitud o semejanza entre ellos obteniendo que analizando que para cada escenario tanto los resultados numéricos como gráficos contenían el mismo modelo. También se evidenciaron muestras en donde a nivel gráfico se obtuvieron dos modelos con el mismo nivel de exactitud y en el numérico solo un modelo, lo cual permite concluir que entre más verificaciones y pruebas se tengan al momento de realizar un análisis mediante modelos los datos generados por el proceso, mayor será la eficacia de la selección del modelo.

Ahora bien, ya obtenidos los modelos seleccionados por escenarios, el paso a seguir según la metodología planteada en la figura 12, es la comprobación de que efectivamente el modelo seleccionado se acopla al comportamiento de las actividades por escenario. Para ello es necesario realizar una evaluación simulando un “tiempo real” del comportamiento del proceso con el modelo incluido.

## **7.2. Generación y análisis de resultados por escenario.**

Una representación gráfica de las principales métricas o KPIs que intervienen en la consecución de las actividades de un proceso se le conoce como Dashboard, y es en ahí en donde se generarán los análisis de la parte final de la metodología predictiva como se muestra en en la figura 13.

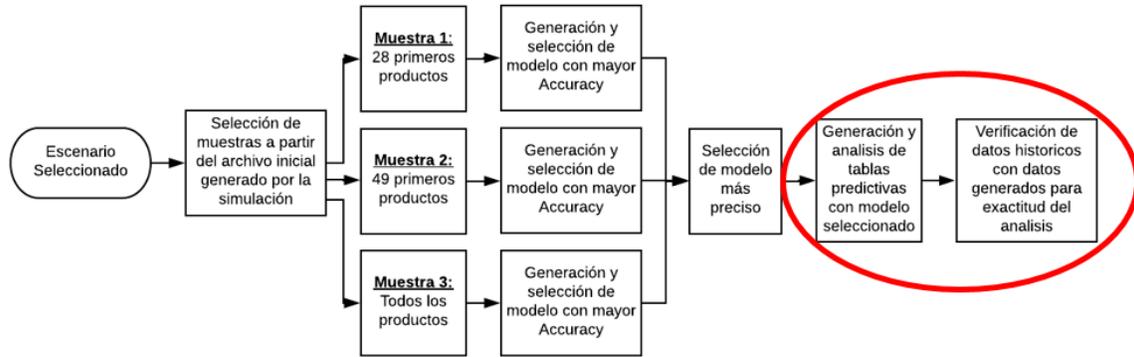


Figura 13. Selección del proceso a analizar en el numeral 6.2. Elaboración propia.

Para la generación de los Dashboards predictivos de los diferentes escenarios, se realizó un proceso de ejecución por pasos, presentado en la figura 14, con el fin de obtener lo resultados más acertados. Se definió una de las muestras como la “Muestra base” la cual sería la que se entrenaría con el modelo elegido, según sea el escenario en el que se esté generando. Luego, esa muestra base fue utilizada para generar los modelos de predicción, posteriormente validados en una “Muestra completa” para así generar predicciones de los datos a manera de tabla de probabilidades o tabla de predicciones. Dichas muestras se definieron así:

- **Muestra base** = Muestra de 28 productos (Muestra 1)
- **Muestra completa** = Muestra de todos los productos (Muestra 3).

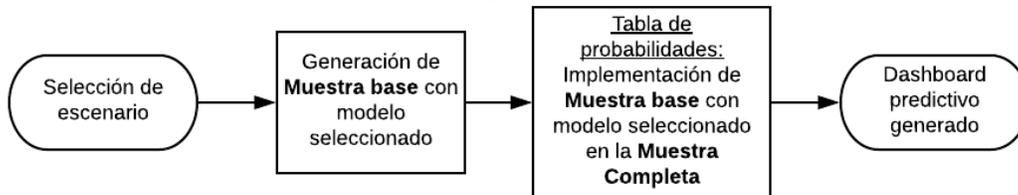


Figura 14. Proceso de generación de Dashboard. Elaboración propia.

Cabe resaltar que, aunque se usó la muestra completa para la validación de los modelos, el total de productos estudiados osciló entre 50 y 70, en otras palabras, se entrenó cada modelo de cada escenario con una muestra de 18 productos (o casos) correspondientes al escenario estudiado y se realizó la validación respectiva con una cantidad de entre 50 y 70 productos tomados de los datos históricos completos (1400 productos). Usualmente, dentro de técnicas de predicción como los árboles de decisión o el análisis discriminante, se deben contrastar los resultados (que para este proyecto son los modelos entrenados) con datos que no han sido usados para la construcción del modelo, con el objetivo de verificar o validar su confiabilidad (Van Der Aalst, 2016), por esta razón es necesario en este punto también realizar la validación correspondiente. Para cada uno de los escenarios estudiados, a continuación, se muestran los resultados de cada validación, a través de un cuadro comparativo de actividades predichas y actividades realmente ejecutadas.

Control Alloc - Cyrille Pach		Actividad predicha						
		Movimiento	Oper.1	Oper.2	Oper.3	Oper.4	Oper.5	Oper.6
Actividad realizada	Movimiento	1499	0	2	0	3	94	0
	Oper. 1	0	203	0	0	0	0	0
	Oper. 2	5	0	73	0	0	0	0
	Oper. 3	36	0	0	6	0	14	0
	Oper. 4	8	0	0	0	36	0	0
	Oper. 5	0	0	0	0	0	56	0
	Oper. 6	78	0	0	0	0	0	0

Potential fields		Actividad predicha						
		Movimiento	Oper.1	Oper.2	Oper.3	Oper.4	Oper.5	Oper.6
Actividad realizada	Movimiento	1375	0	0	0	0	0	0
	Oper. 1	0	164	0	0	0	0	0
	Oper. 2	0	0	63	0	0	0	0
	Oper. 3	0	0	0	45	0	0	0
	Oper. 4	0	0	0	0	35	0	0
	Oper. 5	0	0	0	0	0	45	0
	Oper. 6	61	0	0	0	0	0	0

First Available Machine		Actividad predicha						
		Movimiento	Oper.1	Oper.2	Oper.3	Oper.4	Oper.5	Oper.6
Actividad realizada	Movimiento	626	0	0	0	0	0	0
	Oper. 1	0	0	0	0	0	0	0
	Oper. 2	0	0	0	0	0	0	0
	Oper. 3	0	0	0	0	0	0	0
	Oper. 4	0	0	0	0	0	0	0
	Oper. 5	0	0	0	0	0	0	0
	Oper. 6	0	0	0	0	0	0	0

Randomly		Actividad predicha						
		Movimiento	Oper.1	Oper.2	Oper.3	Oper.4	Oper.5	Oper.6
Actividad realizada	Movimiento	675	0	0	0	0	0	0
	Oper. 1	26	0	0	0	0	0	0
	Oper. 2	0	0	0	0	0	0	0
	Oper. 3	0	0	0	0	0	0	0
	Oper. 4	0	0	0	0	0	0	0
	Oper. 5	1	0	0	0	0	0	0
	Oper. 6	0	0	0	0	0	0	0

**Tabla 9.** Tabla de resumen de predicción de actividades y comparación con datos reales. Elaboración propia.

Nótese que aquellos escenarios que generan mejores resultados al momento del desarrollo de un modelo predictivo corresponden a Control Alloc – Cyrille Pach y Potential fields. Esto coincide con lo observado antes y descrito en la sección 6, donde la baja variabilidad de los tiempos de ciclo, una baja cantidad de variantes y las altas tasas de producción de estos escenarios en comparación otros, sugirieron la conveniencia de la aplicación y desarrollo de un modelo de predicción. Por otra parte, los escenarios First Available Machine y Randomly, no generan predicciones apropiadas. De hecho, los modelos desarrollados para estos escenarios solo predicen en algunos eventos las operaciones de Movimiento y Oper 1. Otras operaciones son omitidas en el modelo debido a la alta variabilidad y aleatoriedad (en el caso de Randomly) de los escenarios en la toma de decisiones, que no permiten encontrar patrones definidos para la generación del modelo.

A pesar de que se obtuvo los mejores modelos con los escenarios Control Alloc – Cyrille Pach y Potential fields, estos escenarios no realizan predicciones sobre la operación 6. Esto puede ser causado por el bajo número de ejecuciones de la actividad en comparación con el resto de las actividades, puesto que, aunque esta actividad debe ser realizada por todos los productos, solo se ejecuta una vez. Adicionalmente, la duración de dicha operación al ser menor, pudo no ser tomada como representativa por el modelo para las predicciones correspondientes, como si lo hace con el resto de las operaciones. Este tipo de fenómenos puede limitar el impacto de la utilización de modelos predictivos en sistemas de manufactura.

La salida de datos de predicción para cada escenario se puede consultar en el archivo de excel anexo titulado "Registros predictivos por escenario" o [AQUÍ](#). En este archivo, además de las salidas de datos predictivos, para cada escenario se muestra una comparación del orden de salida de los datos históricos y los datos predictivos, donde es posible observar el funcionamiento de *Apromore*: mientras la plataforma es alimentada con datos históricos, dichos datos son analizados paulatinamente de acuerdo con el modelo entrenado para arrojar predicciones que en algunos escenarios terminan siendo acertadas. Para una mayor comprensión, a continuación, se explican los elementos más relevantes que conforman los registros predictivos de *Apromore*:

- **Case (Caso):** Número o ID del producto en realización
- **Activity (Actividad):** Operación o actividad realizada por el producto con base en los datos históricos.
- **Elapsed:** Tiempo real transcurrido desde el caso o producto entró al sistema
- **Next Activity:** Probabilidad predicha de ejecución de cada actividad dado modelo entrenado.

A partir de la información hasta aquí mencionada, en la siguiente sección se abordan los resultados y conclusiones.

## 8. Resultados

La minería de procesos, puesta en marcha en este trabajo, se tomó como aquel mecanismo que tiene como objetivo recurrir a los datos generados del simulador seleccionado con el fin de obtener, monitorear y mejorar

el proceso de elaboración de los diferentes productos según el escenario, a través del análisis exhaustivo del registro de los eventos del proceso.

A través de la realización del trabajo se explicó todo lo referente a la parte de obtención y monitoreo de los resultados de cada una de las fases realizadas, ya sea la fase descriptiva o la fase predictiva. Continuando con los objetivos a tener en cuenta, el punto faltante es la explicación de la mejora del proceso según los escenarios seleccionados y aptos para la metodología propuesta.

Los escenarios seleccionados para el desarrollo de este trabajo son escenarios que se componen de trabajos de decisión en cuanto a la siguiente máquina a seleccionar con el fin de cumplir con el proceso de producción. Dichos escenarios se ejecutaron y arrojaron los registros de eventos utilizados para la realización de los análisis respectivos (descriptivos y predictivos). Al obtener estos análisis se pudieron evidenciar dos clasificaciones, los escenarios que se adaptan al análisis de minería de procesos y aquellos que no. Los escenarios de Randomly y FAM obtuvieron resultados poco alentadores para esta metodología debido a su baja tasa de producción y su alta desviación estándar, es decir que, dichos modelos poseen un porcentaje alto de variabilidad en su proceso siendo nula su predicción solo basado en sus datos históricos. Por otro lado, los escenarios CEP y Cyrille poseen una alta tasa de producción y una baja desviación estándar, es decir que, dichos modelos poseen un porcentaje bajo de variabilidad en su proceso siendo estos los escenarios aptos para la metodología propuesta.

La aplicación de la minería de procesos permitió de la misma forma, identificar variantes del proceso y recursos limitantes. Aunque dichos factores están directamente ligados con la configuración de las operaciones, productos y órdenes de producción, se verificó la utilidad de la minería de procesos en un contexto de producción. En el caso particular, fue posible detectar que las máquinas M2, M3 y M7 son recursos limitantes, debido a que la totalidad de los productos deben utilizar alguno de esos recursos. Así mismo, se detectó que la actividad que genera retrasos, está relacionada con los movimientos entre máquinas, y especialmente el movimiento de producto desde su ingreso al sistema hasta la máquina que ejecutará la primera operación. Estos resultados permiten la toma de decisiones, por ejemplo se podría pensar en habilitar más recursos para la realización de la operación 1 y/o 2 que son las que realizan las máquinas M2, M3, y M7 y en términos generales aumentan el tiempo de ciclo

Con base en los resultados obtenidos en *Apromore*, se realizó un dashboard en el programa Excel con los datos de cada escenario. Esto con el fin de ampliar el análisis de la información con ayuda de tablas dinámicas y gráficos. Dicho dashboard se encuentra en la hoja de Excel anexa titulada “Dashboard predictivo” o [AQUÍ](#). Adicionalmente, se realizó un instructivo para un mejor entendimiento del Dashboard y la información que presenta, este archivo se encuentra en el anexo titulado “Instructivo Funcionamiento Dashboard”.

### **8.1 Evaluación del Impacto de la metodología**

Aunque el alcance del presente proyecto, sólo permite la visualización de una interfaz con predicciones de futuras actividades que realizará cada producto, dicha información utilizada en tiempo real permitiría el desarrollo de indicadores predictivos de gran valor para la industrial. La implementación de esta metodología, en los escenarios aptos para ello, genera un impacto relevante a la hora de la producción del sistema de manufactura flexible utilizado. Dado que se desarrolla un entrenamiento de modelos que nos arroja probabilidades y predicciones de la siguiente máquina a utilizar en un proceso, es posible generar una predicción de varios aspectos dentro del proceso:

1. Disponibilidad de recursos por máquina.
2. Tiempo de ciclo
3. Tasa de producción
4. Cuellos de Botella

➤ Disponibilidad de recursos por máquina

El hecho de que en un sistema se logre predecir con anterioridad la cantidad de recursos requeridos, haciendo uso de la metodología propuesta en este trabajo, implicaría una disminución en el nivel de piezas faltantes o tiempos muertos por recarga de recurso en el sistema de producción auto organizado, ya que desde el inicio se

seleccionaría la cantidad puntual de recurso requerido por el periodo de tiempo necesitado. Un ejemplo de ello se puede evidenciar en el escenario de CEP. Al seleccionar dicho escenario se realiza el análisis de descriptivo y predictivo con la metodología propuesta obteniendo que la ruta más probable a realizarse para la generación del producto A es el siguiente:

Oper. 1 → Oper. 2 → Mov. → Oper. 4 → Mov. → Oper. 5 → Oper. 3 → Mov. → Oper. 6

Dado que por necesidades de la producción se requieran hacer 30 productos tipo A en dicho escenario se pueden asignar la cantidad de recursos requeridos por máquina.

➤ Tiempo de ciclo

Por medio de la visualización de las actividades o eventos que tendrá un caso en particular, será posible la estimación del tiempo remanente de un producto en particular. Información de este tipo, permite entre otras cosas planear de mejor forma la producción e incluso generar políticas para el mejor aprovechamiento de los recursos.

➤ Tasa de producción

A partir de los tiempos de ciclo y la cantidad de trabajos en un momento dado, también es posible la generación de un indicador predictivo para la tasa de producción. Dicho indicador, permitirá la identificación de ineficiencias en el sistema con antelación, permitiendo al sistema tomar decisiones autónomas o al dueño del proceso reestructurar de una mejor manera las órdenes de producción, de acuerdo a las necesidades en un momento dado.

➤ Cuellos de botella

Conociendo las actividades u operaciones que cada recurso está en la capacidad de realizar, así como las actividades futuras de los productos en ejecución, es posible desarrollar un indicador que con cierto grado de confiabilidad, muestre recursos que tendrán alto nivel de utilización y puedan ser considerados limitantes del procesos en un momento dado. Esto permitirá junto con los demás indicadores entender de una mejor forma el estado futuro del proceso, permitiendo distribuir mejor las operaciones entre los recursos disponibles, evitando altas cantidades de producto en proceso que en un momento dado puedan afectar el tiempo de ciclo.

Finalmente, el desarrollo del presente trabajo puede ser considerado como una contribución al avance de la temática. Teniendo en cuenta que recientemente pocos artículos han tratado el tema de la aplicación de la minería de procesos en sistemas de manufactura y que ninguno (al menos en la revisión de literatura realizada) ha tratado el monitoreo predictivo de procesos en dichos sistemas de manufactura, el presente documento brinda las bases para la profundización y la continuidad del estudio del tema. Implementación de nuevos modelos de predicción, aplicación de modelos predictivos en industrias específicas, e integración de plataformas (como *celonis*, *apromore* y *nirdizati*) para la generación y predicción de datos, así como el desarrollo de nuevos indicadores a partir de predicciones de actividades, son temas que se proponen como continuidad a lo aquí realizado

## 8. Conclusiones y recomendaciones.

En el análisis descriptivo, al extraer los niveles de utilización de los recursos arrojados por el programa Dashboard de minería de procesos, se puede concluir que, al realizar una redistribución de las actividades asignadas a cada uno, lo cual es posible ya que se trata de un sistema de manufactura flexible, se podría generar una mejora en el desempeño del sistema al reducir tiempos de desplazamiento causados por la espera de aquellos casos que necesitan realizar cierta operación y las máquinas se encuentran ocupadas. Así mismo, otra opción sería habilitar los 7 recursos para que puedan realizar todas las operaciones, de esta manera no existirían recursos limitantes por falta de unidades y exceso de casos pendientes por ejecutar.

Al observar el análisis inicial realizado con los dashboard de minería de procesos, es decir, el análisis descriptivo, se pudo evidenciar que los escenarios Potential fields y Control Alloc-Cyrille Pach son los más opcionados para generar y entrenar modelos de predicción, esto debido a su tasa de producción alta, su baja desviación estándar en el tiempo de ciclo por caso con respecto a los 2 escenarios restantes y la baja cantidad de variantes. Cuando se observa el análisis del Next Activity, realizado con el modelo escogido para cada escenario por su exactitud, se evidencia una relación directa entre lo analizado en la parte descriptiva y la predictiva, ya que los mejores resultados al momento de verificar los datos generados por los dashboards y los

datos históricos, fueron efectivamente los de los escenarios Potential fields y Control Alloc-Cyrille pach. Esto lleva a corroborar que, cuando se realizan análisis de datos ya sea del mismo contexto del presente trabajo u otro, es necesario realizar análisis que puedan compararse entre sí para dar veracidad y confiabilidad a los resultados.

Se puede evidenciar que hay una relación directa entre la cantidad de factores que se tienen en cuenta al momento de determinar un escenario y el desarrollo de un modelo predictivo confiable para un sistema de manufactura similar a este, es decir, que entre más factores se tengan en cuenta dentro de un escenario para definir cuál será la ruta por la que se dirigirá el producto, más confiable será el modelo predictivo. Esto debido a que a mayor número de condiciones haya dentro del escenario, menor será la varianza del tiempo, las tasas de producción serán más altas y el tiempo estándar será menor debido a la reducción de tiempos de desplazamiento entre una máquina y otra, lo que influye de manera importante en el modelo predictivo de tal forma que las probabilidades de que los resultados de este realmente se ejecuten según los datos históricos sean considerablemente altas.

### **Recomendaciones:**

Debido a que los escenarios Potential fields y Control Alloc-Cyrille Pach son aquellos con mejores resultados al realizar los análisis descriptivos y predictivos y cuyos modelos escogidos se aproximan con significativa exactitud a los datos históricos del sistema, se recomienda evaluar la opción de un re diseño de la distribución de la planta de manufactura con base en las variantes con menor tiempo de ciclo de dichos escenarios. Esto se haría con el objetivo de reducir aún más los tiempos de procesamiento, lo cual implica un aumento en la tasa de producción y así, mismo una reducción en las holguras de algunos recursos, haciendo el sistema aún más eficiente.

A partir de los datos obtenidos con respecto a los mejores escenarios para este sistema de manufactura los cuales son Potential fields y Control Alloc-Cyrille Pach, se sugiere que el tiempo de duración promedio de los mismos pueda ser usado para la planeación de la producción, ya que no solo será más acertada en cuanto a los tiempos de producción sino también lo será con respecto a los tiempos de entrega y/o despacho de la misma, ayudando así a mitigar en los gastos propios de la producción, faltantes, entre otros, que pueden derivarse de un cálculo basado en un tiempo cuya exactitud no es lo suficientemente confiable.

Para futuros trabajos, se recomienda la integración del simulador y de *Apromore*, de esta manera mientras el simulador genera datos, *Apromore* genera predicciones en tiempo real. Esto con el propósito de hacer predicciones y análisis de la información de manera más eficiente y eficaz, puesto que se tiene la información en una sola plataforma haciendo más sencillo su manejo. Una propuesta para el output de las predicciones, puede ser el dashboard predictivo presentado en el presente trabajo.

## **5. Glosario**

**Apromore:** Plataforma analítica de procesos de negocios de fuente abierta que combina capacidades de minería de procesos de última generación con funcionalidad avanzada para administrar colecciones de modelos de procesos.<sup>1</sup>

**Nirdizati:** Motor de monitoreo de procesos predictivos basado en web de código abierto para ejecutar procesos.<sup>2</sup>

**NetLogo:** Entorno de modelado programable de agentes múltiples.<sup>3</sup>

**ProM:** Herramienta académica de código abierto para la Minería de Procesos. La cual permite el proceso de descubrimiento, la comprobación de la conformidad, análisis de redes sociales, la minería de organización, la minería de decisión.<sup>4</sup>

---

<sup>1</sup> Recuperado de: [apromore.org](http://apromore.org)

<sup>2</sup> Recuperado de: <http://nirdizati.org/>

<sup>3</sup> Recuperado de: <http://ccl.northwestern.edu/netlogo/>

<sup>4</sup> Recuperado de: [https://www.ecured.cu/Miner%C3%ADa\\_de\\_procesos](https://www.ecured.cu/Miner%C3%ADa_de_procesos)

**Celonis** : Es una herramienta que permite realizar minería de datos con el fin de ofrecer información objetiva, basada en hechos, derivada de registros de eventos reales, que ayudan a auditar, analizar y mejorar procesos existentes al responder preguntas relacionadas con el cumplimiento y relacionadas con el desempeño.<sup>5</sup>

**Caso:** Número de Identificación para cada uno de los productos procesados. Cada producto, corresponde a un caso específico.

**Evento:** Indica el número de operación correspondiente a un producto( caso) procesado.

**Variantes:** Cada uno de los recorridos o caminos particulares de los productos dentro del sistema de manufactura, desde la entrada al sistema, movimientos, ejecución de operaciones y salida

## Referencias

Wilensky, U. 1999. NetLogo. Evanston, IL: Center for Connected Learning and Computer- Based Modeling, Northwestern University. Recuperado de <http://ccl.northwestern.edu/netlogo/>

Aguirre Mayorga, H., & Rincón García, N. (2015). Minería de procesos: desarrollo, aplicaciones y factores críticos. *Cuadernos De Administración*, 28(50), 137-157. Recuperado de <https://doi.org/10.11144/Javeriana.cao28-50.mpda>

Verenich, I. (2016). A general framework for predictive business process monitoring. Recuperado de <https://eprints.qut.edu.au/97539/>

Senderovich, A., Francescomarino, C. D., & Maggi, F. M. (2019). From knowledge-driven to data-driven inter-case feature encoding in predictive process monitoring. *Information Systems*. Recuperado de <https://doi-org.ezproxy.javeriana.edu.co/10.1016/j.is.2019.01.007>

Aalst, W.M., Reijers, H.A., Weijters, A.J., Dongen, B.F., Medeiros, A.K., Song, M., & Verbeek, H.M. (2007). Business process mining: An industrial application. *Inf. Syst.*, 32, 713-732.

Accard, P. (2018). Criticality: How Changes Preserve Stability in Self-Organizing Systems. *Organization studies*, doi:10.1177/0170840618783342

Agudelo L, Escobar, J.. (2007). *Gestión por procesos*. Medellín: los autores.

H. Jalonen and A. Lönnqvist, "Predictive business—Fresh initiative or old wine in a new bottle," *Manage. Decis.*, vol. 47, no. 10, pp. 1595–1609, 2009

Jimenez, J., Zambrano, G., Aguirre S., Trentesaux, D. (2018). Using process - mining for understating the emergence of self - organizing manufacturing systems . Julio, 2018

Mayer, V., Cukier, K. (2013). More. En *Big Data. A Revolution That Will Transform How We Live, Work, and Think* (pp.20-30). Boston, Massachusetts: Houghton Mifflin Harcourt. Recuperado de: <http://catedradatos.com.ar/media/3.-Big-data.-La-revolucion-de-los-datos-masivos-Noema-Spanish-Edition-Viktor-Mayer-Schonberger-Kenneth-Cukier.pdf>

Metzger, A., Leitner, P., Ivanovic, D., Schmieders, E., Franklin, R., Carro, M., Pohl, K. (2015). Comparing and combining predictive business process monitoring techniques. *IEEE Transactions on Cybernetics*, (2), 276. Recuperado de: <http://ezproxy.javeriana.edu.co:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsgeo&AN=edsgcl.406976237&lang=es&site=eds-live>

Pérez, S.. (2015). *Minería de procesos*. Julio, 2018, de Universidad Pontificia Bolivariana. Recuperado de : <https://bit.ly/2RvQsgb>

---

<sup>5</sup> Recuperado de: <https://www.celonis.com/process-mining/what-is-process-mining/>

- Roldán, J., Barrientos, A. (2016). Aplicación de la minería de procesos al análisis y modelado de misiones multi - robot.. Julio, 2018, de Universidad Politécnica de Madrid. Recuperado de: [http://oa.upm.es/43904/1/TFG\\_FELIX\\_DIEZ\\_ARIAS.pdf](http://oa.upm.es/43904/1/TFG_FELIX_DIEZ_ARIAS.pdf)
- Silva Belisario, L., & Pierreval, H. (2013). A conceptual framework for analyzing adaptable and reconfigurable manufacturing systems. In Proceedings of 2013 International Conference on Industrial Engineering and Systems Management. 644–650 Recuperado de: <https://bit.ly/2Qx731O>
- Teinmaa, I., Dumas, M., Rosa, M.L., Maggi, F.M.: Outcome-oriented predictive process monitoring: Review and benchmark. CoRR abs/1707.06766 (2017)
- T. Tolio, Design of Flexible Production Systems. Springer: Berlin, 2009. Recuperado de: <https://bit.ly/2E2sCpN>
- Trentesaux, D., Pach, C., Bekrar, A., Sallez, Y., Berger, T., Bonte, T., Leitão, P. & Barbosa, I. (2013). Benchmarking flexible job-shop scheduling and control systems. agosto 11, 2018, de Pontificia Universidad Javeriana. Recuperado de: <https://bit.ly/2BZRFqI>
- Van Der Aalst, W. (2016). Process Mining: Data Science in Action. Berlín: Springer.
- Verenich, I., Mõškovski, S., Raboczi, S., Dumas, M., La Rosa, M., & Maggi, F. (2018). Predictive Process Monitoring in Apromore. julio 28, 2018, de Queensland University of Technology. Recuperado de: <https://eprints.qut.edu.au/117955/>

## Anexos del Trabajo de Grado 183013

A continuación, se presenta la lista de los anexos presentados en este trabajo de grado.

Nombre	Enlace	Extensión
Anexos	<a href="D:\Anexos\Anexos.pdf">D:\Anexos\Anexos.pdf</a>	pdf
Dashboards Predictivos	<a href="D:\Anexos\Dashboards Predictivos.xlsx">D:\Anexos\Dashboards Predictivos.xlsx</a>	xlsx
Instructivo funcionamiento dashboard	<a href="D:\Anexos\Instructivo funcionamiento dashboard.pdf">D:\Anexos\Instructivo funcionamiento dashboard.pdf</a>	pdf
Sistema de manufactura simulado	<a href="D:\Anexos\Sistema de manufactura simulado.rar">D:\Anexos\Sistema de manufactura simulado.rar</a>	rar
Registros de eventos por escenario	<a href="D:\Anexos\Registros de eventos por escenario.rar">D:\Anexos\Registros de eventos por escenario.rar</a>	rar
Registros predictivos por escenario	<a href="D:\Anexos\Registros predictivos por escenario.xlsx">D:\Anexos\Registros predictivos por escenario.xlsx</a>	xlsx

Este documento fue creado automáticamente. Para cualquier comentario comunicarse con la coordinación de trabajos de grado.