

MODELO PREDICTIVO DESDE ENTIDADES CONTRATANTES DE PROVEEDORES PARA CALL CENTER CON IMPLEMENTACIÓN DE BOTS BASADO EN DATOS DADOS POR EL ESTÁNDAR COPC

Adrian Felipe Pinzón Hurtado

Master de Ciencias de Datos

Estadística Avanzada / Minería de Datos / Data Mining

Nombre Consultor/a Eduardo Gil Blasco

Nombre Profesor/a responsable de la asignatura: Albert Solé Ribalta

Fecha Entrega 24/06/2022



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada a [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)

A) Creative Commons:



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada a [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual [3.0 España de Creative Commons](#)



Esta obra está sujeta a una licencia de Reconocimiento [3.0 España de Creative Commons](#)

B) GNU Free Documentation License (GNU FDL)

Copyright © AÑO TU-NOMBRE.

Permission is granted to copy, distribute and/or modify this document under the terms of the

GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (el autor/a)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Modelo Predictivo para Call Center con funcionamiento de bots basado en Datos Dados por el Estándar COPC</i>
Nombre del autor:	<i>Adrian Felipe Pinzón Hurtado</i>
Nombre del consultor/a:	<i>Eduard Gil Blasco</i>
Nombre del PRA:	<i>Albert Solé Ribalta</i>
Fecha de entrega (mm/aaaa):	24/2021
Titulación::	<i>Master en Ciencias de Datos</i>
Área del Trabajo Final:	<i>Estadística Avanzada / Minería de Datos</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>COPC</i> <i>Call Center</i> <i>Modelo predictivo call</i> <i>Bots</i>
Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i>	
<p>Las modelos de forecasting en call center han tenido una perspectiva desde el prestador de servicios para predecir cuántos agentes humanos poner en línea, sin tener en cuenta las compañías contratantes de estos servicios (clientes), que muchas veces no tienen dentro de su estrategia un análisis de datos para predecir contactos para las nuevas tendencias en bots y mejorar el servicio.</p> <p>La mayoría recibe información de <i>standard</i> COPC, que incluye llamadas atendidas, llamadas colgadas, llamadas abandonadas, tiempo de atención, AHT, TSF, CSF, entre otros. Estos datos que son tan importantes, han sido muy poco utilizados por las compañías contratantes para tener una perspectiva más amplia por la tendencia de los bots, y que es necesario para el futuro comportamiento de los call center (menos humanos, más contactos para los bots).</p> <p>La metodología aplicada será un modelo de regresión logística con el fin de dar un abanico de posibilidades al analista de infraestructura encargado del correcto funcionamiento de los bots.</p>	

El resultado del trabajo será un nuevo modelo y las distintas recomendaciones para tener la mejor predicción que permita garantizar la alta disponibilidad de los bots.

Abstract (in English, 250 words or less):

The call center forecasting models have had a perspective from the service provider predicting how many human agents have to be on line, without taking into account the companies that contract these services, which often do not have data analysis to predict contacts and improve the service within their strategy counting with the new trends of bots in call centers.

Most receive COPC standard information, which includes metrics like time, Calls Answered, Svc Level (% answered w/i 60 sec), Avg Speed Answer (sec), Transferred Calls % which are so important, have been little used by contractors to have a broader perspective of the possible behavior of their call center.

However, the bots in the market will replace humans in the near future, so the strategy has to be focused on how to predict contacts in order to guarantee the availability of the bot infrastructure.

The methodology applied will be different prediction models such as logistic regression, SVA and decision tree in order to provide a range of possibilities.

The results of the work will be the best model and the different recommendations to have the best prediction and applying the best technology for bots.

Índice

1. Introducción	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	1
1.3 Enfoque y método seguido	1
1.4 Planificación del Trabajo	1
1.5 Breve sumario de productos obtenidos	1
1.6 Breve descripción de los otros capítulos de la memoria	1
2. Resto de capítulos	2
3. Conclusiones	3
4. Glosario	4
5. Bibliografía	5
6. Anexos	6

Lista de figuras

No se encuentran elementos de tabla de ilustraciones.

1. Introducción

1.1 Contexto y justificación del Trabajo

El punto de partida del trabajo son datos suministrados en distintos call center según la metodología COPC. Es un tema relevante para todas las compañías que reciben estos datos por parte de un proveedor del servicio de call center y necesitan revisar a futuro el comportamiento para mejorar el servicio y optimizar los costos.

En el momento la problemática es resuelta en parte por modelos ofrecidos en el mercado como Erlang y ARIMA, los cuales exigen datos que en muchos casos no son suministrados a nivel contractual por los prestadores de servicio de call center.

La principal motivación es que el call center es el punto de contacto de los clientes con las empresas para dar un servicio y es vital tener un panorama de servicios versus costo. Las inversiones en call centers son muy grandes y pueden oscilar entre USD 170000 y USD 200000 para 50000 a 60000 contactos en promedio mensual. Esto incluye recurso humano, infraestructura, administración y locación.

Aunque actualmente uno de los puntos más caros en el recurso humano, este ha sido paulatinamente reemplazado por bots gracias a la cuarta revolución industrial. Por esta razón, los indicadores de servicio (COPC) y los costos facturados por los proveedores, son las partes más importantes para medir y tomar decisiones en cuanto al servicio y el precio dispuesto a pagar. La optimización puede suponer un gran ahorro para la empresa además de asegurar la calidad del servicio para los clientes.

El resultado a obtener es un modelo más robusto hecho en R o *python*, maleable a datos disponibles y estándar para cualquier entidad que terceriza su call center, que ya en sí cuenta con una metodología (COPC), y que arroje una tabla de precios con datos de buena precisión y mundialmente aplicados que serán la base para el modelo en contrataciones a futuro.

1.2 Objetivos del Trabajo

Objetivo general

Obtener un modelo predictivo de precios para call center con implementación de bots y datos COPC de entrada.

Objetivos específicos

- Investigar los modelos actuales para call center y tipificar sus falencias para la norma COPC desde la perspectiva cliente contratante.
- Testear distintos datasets de call center.
- Testear nuevos modelos de predicción.
- Escoger el mejor modelo.

1.3 Enfoque y método seguido

Desarrollar un producto nuevo que supla las necesidades de compañías que utilizan servicios de call center para sus distintos servicios.

Esta estrategia es importante para el logro de objetivos ya que en el mercado no se encuentra un producto similar acorde con cada necesidad.

El método seguido fue la entrevista a un CEO de Call Center y a un BPO quienes expusieron los principales desafíos para lograr dos objetivos principales: mejorar el servicio de call center al mejor precio desde un modelo predictivo y poder tener un máximo control del proveedor.

Se utilizará el lenguaje R, librerías ggplot y un modelo de regresión logística múltiple.

La metodología utilizada será agile, donde se tendrán sprints y una entrega mínima viable en primera instancia la cual se irá mejorando hasta la fecha final propuesta en el diagrama Gantt.

1.4 Planificación del Trabajo

Los principales datasets a utilizar están en *kaggle* y en Bancolombia.

En cuanto al conocimiento, se aplicarán conceptos vistos en estadística avanzada, *data mining* y modelos avanzados de minería de datos.

Diagrama de Gantt:

Diagrama de Gantt				Febrero				Marzo				Abril				Mayo				Junio			
Actividades	Inicio	Fin	Días	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Hito 1: Análisis de Mercado de Modelos para Call Center	28/02/2022	07/03/2022	4																				
Investigación modelos predictivos para call center	28/02/2022	03/03/2022	3																				
Investigación standard COPC y métricas	04/03/2022	05/03/2022	1																				
Conclusiones	06/03/2022	06/03/2022	0																				
Gestión de datos	07/03/2022	07/03/2022	0																				
Hito 2: Exploración de datos	08/03/2022	04/04/2022	23																				
Exploración de datasets con datos COPC	08/03/2022	14/03/2022	6																				
Preprocesado de datos	15/03/2022	25/03/2022	10																				
Análisis estadístico inicial	26/03/2022	28/03/2022	2																				
Limpieza de datos	29/03/2022	01/04/2022	3																				
Conclusiones iniciales	02/04/2022	04/04/2022	2																				
Hito 3: Modelo Predictivo	05/04/2022	03/05/2022	24																				
Exploración modelo predictivo regresión logística	05/04/2022	07/04/2022	2																				
Exploración modelo predictivo RL múltiple	08/04/2022	15/04/2022	7																				
Exploración modelo predictivo Random forest	16/04/2022	21/04/2022	5																				
Exploración Predicciones más Acertadas	22/04/2022	30/04/2022	8																				
Conclusiones	01/05/2022	03/05/2022	2																				
Hito 4: Documentación y Presentación del proyecto	04/05/2022	24/06/2022	48																				
Preparación y redacción de la memoria	04/05/2022	28/05/2022	24																				
Revisión de la memoria con el consultor	29/05/2022	05/06/2022	7																				
Preparación de la defensa del proyecto	06/06/2022	12/06/2022	6																				
Defensa pública	13/06/2022	24/06/2022	11																				

1.5 Breve resumen de productos obtenidos

- 1.5.1. Presentación y problemática.
- 1.5.2. Descripción del dataset.
- 1.5.3. Preprocesamiento y limpieza de datos.
 - 1.5.3.1. Valores Perdidos.
 - 1.5.3.2. Valores Extremos
- 1.5.4. Modelo de predicción de costos con distintas opciones de datos entrantes.
- 1.5.5. Conclusiones.

1.6 Breve descripción de los otros capítulos de la memoria

- Conceptos Básicos Normas COPC.

- Análisis de mercado modelos para call center desde perspectiva cliente/proveedor.
- Falencias actuales modelos para call center desde perspectiva cliente.
- Modelo predictivo más cercano.

2. Estado del Arte

2.1. Antecedentes

En el mercado, el estándar mundialmente aplicado para la administración y medición de call center (COPC), tiene como principales indicadores y conceptos los siguientes:

Concepto	Definición
AHT (Average Handle Time)	Es el tiempo de conversación esperado en promedio, expresado en minutos, y también discriminado por los canales del call center (teléfono, chat y web/mail).
Contactos atendidos	Contactos que fueron contestados en el call center
Contactos ofrecidos	Contactos que entraron al call center (contestados, no contestados y abandonados).
Línea base	Línea Base: es el número de contactos que se toman como base y que se esperan atender ese mes, por lo cual a partir de esta base se factura anticipadamente.
Umbral	Segundos o tiempo acordado para contestar un contacto desde el momento que el usuario/cliente se comunica
Contestados antes del Umbral	Contactos contestados antes del umbral definido. Ejemplo: si se define 25 segundos, los contactos contestados antes del umbral es el total de contactos que se contestaron antes de los 25 segundos
TSF (Telephone Service Factor)	Factor de servicio telefónico, corresponde a la división de los contactos antes del umbral / total de contactos atendidos para el medio de teléfono
CSF (Chat Service Factor)	Factor de servicio en chat, corresponde a la división de los contactos antes del umbral / total de contactos atendidos para el medio de chat

A pesar de que los consumidores prefieren aún el canal telefónico (61%) seguido del email (60%) y chat (57%), están aumentando las preferencias por consultas en bases de conocimiento (51%) y autoreporte (51%).¹

Estos indicadores están en estos momentos siendo reevaluados por la nueva tendencia de *bots*, los cuales en un futuro reemplazarían totalmente a los agentes humanos. Esto depende mucho de la evolución de la inteligencia artificial.

Aunque se están realizando avances importantes, los call centers como estrategia están manejando los canales humanos en paralelo con los chatbots de una manera discreta y controlada para no afectar el servicio. Las tareas que se están migrando inicialmente a estos chatbots son sencillas, de fácil respuesta y sin mucho crecimiento en el árbol de decisión. Es por tanto el medio chat, el primero que está surgiendo en esta tendencia.

Para el medio de mayor preferencia hasta ahora, como es el telefónico, los *bots* requieren de una alta infraestructura, que aún se está testeando, es cierto que el mercado ya ofrece este canal pero con una gran inversión no sólo en infraestructura, sino también IA y escalabilidad. Su dependencia en la red telefónica, la multicanalidad y los algoritmos de IA aplicada, aún no están tan maduros para ser asumidos como una solución real. Sin embargo, como es un canal ampliamente utilizado por los clientes, se debe tomar muy en cuenta para esta tendencia de bots y esto implica una predicción telefónica que ya no estará enfocada en número de agentes, sino en cuántas llamadas simultáneas podría recibir para así garantizar la alta disponibilidad en infraestructura al mejor coste.

2.2. Modelos económicos para contratar Call Centers

Las contrataciones de servicios de call center, han tenido en su historia modelos de cobro definidos por uno de estos tres aspectos:

- a) Cobro por línea base (rango máximo de contactos que esperan recibir en un mes).
- b) Cobro por puestos de trabajo (PPST): número de puestos que se cobran para atender de cada puesto tanto llamadas como visitas de soporte en sitio (llamados técnicos en campo).
- c) Cobro por número de asesores conectados (en tiempo).

Con las nuevas tendencias en *bots*, el tercer modelo tiende a desaparecer. En cuanto al modelo b), la contratación ha demostrado que favorece más al proveedor que la empresa contratante, porque cada mes no todos los puestos de trabajo contactan al call center.

En el presente trabajo se tomará el modelo a) ya que es el más probable de aplicar en el futuro.

¹ <https://callminer.com/blog/call-center-statistics-you-should-know>

Para este modelo a), los costos mensuales promedios se encuentran en COP 671.270.279 para un promedio de Línea base de 52321 contactos atendidos al mes. Esto es lo que ofrece un proveedor de servicios de alta calidad.

2.3. Modelos Estadísticos Actuales

Los modelos ampliamente ofrecidos por el mercado como Erlang y ARIMA, no cumplen las expectativas desde la compañía contratante, pues su enfoque es más hacia la optimización, además de no tener en cuenta la nueva tendencia en *bots*.

Otra falencia de los modelos, es que no cuentan con los datos necesarios que en muchos casos no suministra el proveedor de servicios, como es la carga de trabajo, o número de agentes (exigido por el modelo Erlang), si se ha contratado los modelos económicos a) y b) planteados anteriormente.

En el caso de ARIMA, la complejidad del modelo, que utiliza series de tiempo, permite una perspectiva para predecir tendencias futuras. Un modelo estadístico es autorregresivo si predice valores futuros basados en valores pasados. ARIMA utiliza promedios móviles rezagados para suavizar los datos de series de tiempo. Los parámetros necesarios para un modelo ARIMA, no están cubiertos en los datos COPC, como por ejemplo los contactos ofrecidos, los contactos contestados, los contactos contestados antes del umbral y el AHT. .

Es por esto importante resaltar que los modelos predictivos actuales han tomado “el número de agentes” como una variable que debe ser reconocida en el modelo, algo que en poco tiempo será innecesario por la llegada de los *bots*.

Por esta razón, el modelo predictivo que se propone, se fundamenta principalmente en tomar datos COPC y rangos de precios actuales para poder tomar decisiones en un RFP (Request for Proposal), es decir, licitaciones con distintos proveedores que puedan dar un precio dispuesto a pagar de acuerdo a las distintas variables entregadas.

Resto de capítulos

- Conceptos Básicos Normas COPC
- Análisis de mercado modelos para call center desde perspectiva cliente y proveedor.
- Falencias actuales modelos para call center desde perspectiva cliente.
- Modelo predictivo más eficaz y robusto.
- Repositorio con data y código:
https://github.com/adrife20/PEC3_TFM_adpinzon

3. Conclusiones

Las variables que se han tomado para el modelo de precios, son las ampliamente utilizadas para la mayoría de modelos de cobro: línea base, AHT de teléfono, AHT de chat, número de masivos (o servicios nuevos), contactos atendidos y contado atendidos en chat.

Los modelos indican que la variable de # de masivos afecta el resultado final de la línea base predicha.

El AHT también afecta considerablemente la línea base y por lo tanto el costo.

Se toma el modelo 7 para la nueva lista de precios con un p-value: $3.774e-12$, mostró una distribución normal y su precisión es aceptable.

Aunque la data muestra algunas variables altamente correlacionadas que se podrían eliminar, como el AHT Chat, este es necesario para el modelo de predicción. La data histórica no mostró esto, pero en la práctica todo call center se afecta por su tiempo de conversación. Por ejemplo, con un tiempo de conversación de 10 minutos se atiende una llamada, pero si el asesor demora 20 minutos con una llamada cuando debería atender dos en este tiempo, pues ahí se confirma que es bastante importante, por lo que se deben tomar en cuenta otras variables.

Una de las variables que más afecta el AHT es un evento masivo (no entregado), que no se puede predecir, pero que ocasiona que el asesor se demore más por un tema que no tiene entregado o no sabe qué hacer. La data recogida mostró la relevancia de los eventos masivos.

Los contactos por bots estaban relacionados negativamente (-0.73) con la gestión indirecta. Esto se interpreta como que entre más contactos atienda el bot, menos contactos van a llegar a un asesor para atender indirectamente.

En la tabla de precios predichos no se tuvo en cuenta estos contactos de bots, por tratarse de un medio bastante barato que no afecta en gran medida el costo. Sin embargo, entre más contactos haya en este canal, los costos se disminuyen, pero esto aplica para el proveedor de servicios, mas no para el cliente contratante si no lo tiene implementado como core.

Es también bastante importante resaltar que aunque se podría aumentar la afectividad y precisión del modelo, se habría tenido que sacrificar importantes variables como el AHT o los contactos atendidos.

4. Glosario

COPC: norma técnica Customer Operations Performance **Center (Copc)** es la que fija los modelos de gestión que deben cumplir estas entidades para ofrecer un servicio excelente.

BPO: Business Process Outsourcing (BPO), el sector de tercerización de procesos de negocio, se entiende como la delegación de uno o más procesos de negocio, intensivos en el uso de tecnologías de la Información, a un proveedor externo, quien a su vez posee, administra y gerencia los procesos seleccionados, basado en métricas definidas y medibles.

ERLANG: Modelo creado por el pionero de la teoría de colas, Agner Krarup Erlang. Se utiliza en la modelización del tráfico telefónico para determinar el número de personal necesario en función de la carga de trabajo prevista para las llamadas entrantes, junto con otras variables como el nivel de servicio deseado (porcentaje de llamadas atendidas en un número determinado de segundos).

ARIMA: Autoregressive Integrated Moving Average. Modelo que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro.

RFP: (Request for Proposal). Solicitud de licitación para un nuevo contrato.

ITIL: Es la Biblioteca de Infraestructura de Tecnologías de Información . Es un conjunto de conceptos y buenas prácticas usadas para la gestión de servicios de tecnologías de la información, el desarrollo de tecnologías de la información y las operaciones relacionadas con la misma en general.

5. Bibliografía

Cecilie Bertinussen Nordheim, Asbjørn Følstad, Cato Alexander Bjørkli, "An Initial Model of Trust in Chatbots for Customer Service" Interacting with Computers, Volume 31, Issue 3, May 2019, Pages 317–335, <https://doi.org/10.1093/iwc/iwz022>

Cichosz Pawel, Data Mining Algorithms, John Wiley Sons Ltd, Chichester United Kingdom, 2015.

Gironés Jordi; Casas Jordi; Minguillon Julia, Minería de Datos Modelos, 2021

Subirats, L., Pérez, D., Calvo, M. Introducción al ciclo de vida de los datos. Editorial UOC. (2019).

Montserrat Guillén Estany, Modelización predictiva: introducción a los modelos lineales generalizados, FUOC PID_00247912.

Raschka Sebastian, Python Machine Learning, Packt Publishing, ProQuest Ebook Central, Birmingham, 2015.

Vegas Lozano Esteban, Preprocesamiento de Datos, PID_0024790

<https://www.kaggle.com/anirbank/callcenteroptimization/data>, febrero 2022

<https://www.kaggle.com/cityofoakland/oakland-call-center-public-work-service-requests>, febrero 2022

<https://www.copc.com/copc-standards/>, febrero 2022.

http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2077-33231997000100005 febrero 2022.

6. Anexos

N/A