

**iic**  
instituto  
de ingeniería  
del conocimiento



# Comunicación Clara

*Resumen ejecutivo*

IIC - Social Business Analytics

30 de Octubre de 2020



[www.iic.uam.es](http://www.iic.uam.es)



**Instituto de Ingeniería  
del Conocimiento**

C/ Francisco Tomás y  
Valiente, nº 11  
Escuela Politécnica  
Superior (EPS),  
Edificio B, 5ª planta  
Universidad Autónoma de  
Madrid (UAM).  
28049 Cantoblanco,  
Madrid

Tel: +34 91 378 50 00

## Resumen ejecutivo

---

- El objetivo del proyecto Ayuntamiento de Madrid - Comunicación Clara es analizar la calidad de las respuestas emitidas por el Ayuntamiento de Madrid a las solicitudes que efectúan los ciudadanos.
- Las respuestas emitidas por el Ayuntamiento se evalúan a partir de dos ejes de calidad: el eje de claridad y el eje de personalización. Cada uno de estos ejes está compuesto por métricas lingüísticas de evaluación de la calidad.
- Las métricas son herramientas de evaluación de aspectos lingüísticos objetivos, medibles y cuantificables en los textos. La selección de las métricas de evaluación de calidad viene condicionada por los recursos disponibles: características del corpus, capacidad de implementación y tiempo.
- El eje de claridad está compuesto por 9 métricas lingüísticas de claridad basadas en elementos que aparecen en los textos (tecnicismos, signos de puntuación, estructuras pasivas...). Estas métricas constituyen la fuente única de información para el modelo de aprendizaje automático de claridad. El modelo de aprendizaje automático de claridad predice la probabilidad que tiene un texto nuevo de ser claro o no claro.
- El eje de personalización está compuesto por 5 métricas lingüísticas de personalización basadas en elementos que aparecen en las respuestas (títulos indeterminados) o elementos que relacionan la solicitud y la respuesta (coincidencia de lemas o entidades nombradas). Estas métricas constituyen la fuente única de información para el modelo de aprendizaje automático de personalización. El modelo de aprendizaje automático de personalización predice la probabilidad que tiene un par de solicitud-respuesta nuevo de ser personalizado o no personalizado.
- El punto de partida es un corpus de 198.889 muestras de ámbito administrativo.
- Las 9 métricas de claridad aportan información relevante de manera individual: permiten representar y analizar el estado del corpus en base a un aspecto lingüístico evaluable de claridad concreto. Conjuntamente, se ha demostrado que todas ellas aportan información al modelo de aprendizaje automático aunque en diferente grado: la puntuación global de claridad se calcula a través de pesos asignados a cada una de las métricas de claridad.
- Las 5 métricas de personalización aportan información relevante de manera individual: permiten representar y analizar el estado del corpus en base a un aspecto lingüístico evaluable de personalización concreto. Conjuntamente, se ha demostrado que todas ellas aportan información al modelo de aprendizaje automático aunque en diferente grado: la puntuación

global de personalización se calcula a través de pesos asignados a cada una de las métricas de personalización.

- El acierto de los modelos de claridad (0,998) y personalización (1) se concentra en los casos extremos: los textos con probabilidades muy altas y muy bajas de ser claros o no claros y personalizados o no personalizados. El acierto de los modelos refleja la capacidad de adaptación de aprendizaje: ambos modelos han aprendido con éxito las métricas y los pesos asociados a ellas.
- El modelo de claridad recibe como entrada el texto respuesta etiquetado como claro o no claro en base a la puntuación global de las métricas de claridad. Con estos datos, el modelo proporciona predicciones de claridad sobre textos nuevos.
- El modelo de personalización recibe como entrada los textos de solicitud y respuesta etiquetados como personalizado o no personalizado en base a la puntuación global de las métricas de personalización. Con estos datos, el modelo proporciona predicciones de personalización sobre pares solicitud-respuesta nuevos.
- Hay disponible una API REST para realizar pruebas texto a texto o de forma masiva. Al final de este informe se proporciona un manual de usuario con las instrucciones necesarias.
- La API funciona con textos en español, similares en extensión y dominio a los textos que componen el corpus proporcionado por el Ayuntamiento de Madrid.
- La manera más eficaz de mejorar el proceso de aprendizaje automático a futuro es contar con un corpus anotado en el que los ciudadanos expresen si las respuestas recibidas son claras o personalizadas.

[www.iic.uam.es](http://www.iic.uam.es)



*Todos los derechos reservados*



**Instituto de Ingeniería del Conocimiento**

C/ Francisco Tomás y Valiente, nº 11  
Escuela Politécnica Superior (EPS), Edificio B, 5ª planta  
Universidad Autónoma de Madrid (UAM).  
28049 Cantoblanco, Madrid

T. (+34) 91 497 23 23

