

AVANCE 6.

“Conclusiones clave”

PROYECTO INTEGRADOR (GPO 10)



PROFESORES TITULARES

DRA. GRETTEL BARCELÓ ALONSO

DR. LUIS EDUARDO FALCÓN MORALES.

EQUIPO 31

JUAN CARLOS VILLAMIL ROJAS A01794003

MATEO CRUZ LANCHERO A01793882

ANDREA MARGARITA OSORIO GONZÁLEZ A01104776

- A. En este avance, se busca fundamentar si es viable implementar el modelo, evaluando su adecuación a los criterios de éxito previamente establecidos y considerando los resultados obtenidos durante la etapa de modelado. En cualquier caso, será importante también incluir los siguientes cuestionamientos:**

¿El rendimiento del modelo es lo suficientemente bueno para su implementación en producción?

Se obtuvieron puntuaciones ROUGE significativamente mejores que el baseline, podemos considerar que el rendimiento es suficientemente bueno para una implementación inicial en producción como parte de la primera iteración realizada.

Estos resultados indican una mejora sustancial sobre el modelo baseline, la cual puede ser tomada como referente para próximas iteraciones de evaluación y mejora.

Comparando los resultados obtenidos con los criterios de éxito definidos en la Fase 0:

1. Rendimiento del modelo:
 - Meta inicial: Lograr un ROUGE-1 > 0.3 y ROUGE-L > 0.25
 - Resultado: El modelo Claude-v1 alcanzó ROUGE-1 de 0.55 y ROUGE-L de 0.50, superando las metas establecidas.
2. Tiempo de generación:
 - Meta inicial: Resúmenes generados en < 2 minutos por transcripción
 - Resultado: Tiempo promedio de generación de 1.5 minutos, cumpliendo el objetivo.
3. Longitud del resumen:
 - Meta inicial: Resúmenes con longitud del 10-15% del texto original
 - Resultado: Longitud promedio del 12%, dentro del rango deseado.

El modelo Claude-v1 supera las metas establecidas y muestra un rendimiento consistente en diferentes tipos de reuniones.

¿Existe margen para mejorar aún más el rendimiento?

Sí, existe margen para mejorar. Recomendamos:

- Fine-tuning con datos específicos de SISTICA
- Fine-tuning con formatos de video diferentes y más semejantes a las que desarrolla SISTICA
- Implementación de post-procesamiento para mejorar coherencia
- Desarrollo de un sistema de aprendizaje continuo basado en feedback de usuarios

¿Cuáles serían las recomendaciones clave para poder implementar la solución?

Recomendaciones para implementación:

1. Establecer un pipeline de datos robusto en la nube para procesar transcripciones.
2. Implementar un sistema de monitoreo en tiempo real del rendimiento del modelo.

3. Desarrollar una interfaz de usuario intuitiva para acceder y editar resúmenes.
4. Establecer un proceso de revisión humana para casos críticos o de baja confianza.

¿Qué tareas / procedimientos son accionables para las partes interesadas (stakeholders)?

Accionables para stakeholders:

1. Para el equipo de gestión:
 - Definir KPIs específicos para la calidad de los resúmenes (ej. precisión > 90%, tiempo ahorrado > 30 min/reunión)
 - Revisar y aprobar mensualmente informes de rendimiento del sistema
2. Para usuarios finales (empleados):
 - Proporcionar feedback semanal sobre la precisión y utilidad de los resúmenes generados
 - Participar en sesiones trimestrales de capacitación sobre el uso efectivo del sistema
3. Para el equipo de TI:
 - Realizar mantenimiento semanal del sistema y actualizaciones mensuales del modelo
 - Generar informes mensuales de rendimiento y uso del sistema
4. Para el equipo legal:
 - Revisar trimestralmente el cumplimiento del sistema con las normativas de protección de datos
 - Actualizar anualmente las políticas de uso y privacidad relacionadas con el sistema

B. Realizar un análisis detallado para determinar la plataforma de servicios en la nube más adecuada para implementar la solución de ML. Para ello deberán: Evaluar las características y ventajas de los principales proveedores como: AWS, Azure, GCP e IBM Watson.

Considerar factores como la facilidad de uso, la escalabilidad, los servicios específicos ofrecidos, los costos asociados y cualquier otro aspecto relevante. Proporcionar una justificación fundamentada para la elección del proveedor que consideren más apropiado.

Comparativa entre los 2 proveedores más óptimos a ser implementados por SISTICA, de acuerdo a los servicios que actualmente tiene:

1. Microsoft Azure:
 - Integración: Excelente integración con sistemas existentes de Microsoft 365
 - Servicios ML: Azure Cognitive Services ofrece potentes capacidades de NLP
 - Seguridad: Cumple con GDPR y ofrece encriptación robusta
 - Costos: Flexible, con opciones de optimización como instancias reservadas

- Facilidad de uso: Interfaz intuitiva, adecuada para equipos con experiencia limitada en ML
- 2. Amazon Web Services (AWS):
 - Integración: Requiere más trabajo para integrarse con sistemas Microsoft
 - Servicios ML: Amplia gama de servicios a través de SageMaker
 - Seguridad: Cumple con GDPR y ofrece opciones avanzadas de seguridad
 - Costos: Generalmente más económico para cargas de trabajo variables
 - Facilidad de uso: Curva de aprendizaje más pronunciada, mejor para equipos experimentados

Selección y argumentación:

Recomendamos Microsoft Azure por las siguientes razones:

1. Integración superior con la infraestructura existente de SISTICA (Microsoft 365 y Slack).
2. Azure Cognitive Services proporciona herramientas de NLP listas para usar, reduciendo el tiempo de desarrollo.
3. La interfaz más amigable se adapta mejor al nivel de experiencia del equipo de TI de SISTICA.
4. Los costos, aunque potencialmente más altos que AWS, se justifican por la facilidad de integración y uso.
5. Las capacidades de seguridad y cumplimiento de Azure se alinean perfectamente con los requisitos de SISTICA.