**Guía3. Informe final Proyecto APT**

**Asignatura Capstone**

|  |
| --- |
| **1. Informe final Proyecto APT** |
| El objetivo de este informe es que describas los aspectos más relevantes de tu Proyecto APT. Es importante que fundamentes las decisiones que tuviste que tomar a lo largo del proceso.  A continuación, encontrarás distintos campos que deberás completar con la información solicitada, los que dan cuenta del resumen de tu proyecto APT y sus principales resultados. |

|  |  |
| --- | --- |
| Nombre del proyecto | **Agente AI Contable** |
| Área (s) de desempeño(s) | * *Ingeniería de Software* * *Inteligencia Artificial y Machine Learning* * *Automatización de Procesos (RPA)* * *Cloud Computing y DevOps* |
| Competencias | * *Desarrollo de software con metodologías ágiles* * *Diseño e implementación de arquitecturas cloud escalables* * *Aplicación de técnicas de Machine Learning para clasificación de datos* * *Integración de sistemas mediante APIs y automatización* * *Gestión de proyectos tecnológicos con enfoque de producto* * *Testing, calidad de software y prácticas DevOps* |

|  |  |
| --- | --- |
| **Contenidos del informe final** | |
| 1. Relevancia del proyecto APT | * El proyecto aborda la ineficiencia operativa en procesos contables de pequeñas y medianas empresas en Perú. Actualmente, los contadores dedican entre 6-8 horas semanales a tareas manuales repetitivas: descargar facturas desde el portal de SUNAT (Superintendencia Nacional de Aduanas y de Administración Tributaria), clasificarlas manualmente según su naturaleza contable (gastos, activos fijos, insumos, etc.), y registrarlas una por una en sistemas contables como CONCAR CB. * ¿Por qué es relevante este tema para el campo laboral de tu carrera? * país: Perú   Sector: empresas medianas y pequeñas con obligaciones tributarias formales.  Sistema objetivo: usuarios de CONCAR CB (software contable líder en Perú)  Portal tributario: Sunat (equivalente al SII chileno)     * Contadores y estudios contables: Profesionales que dedican 30-40% de su tiempo a tareas de bajo valor agregado * Empresas PyME: Organizaciones que pagan entre $500-$1,500 mensuales por servicios contables básicos * ***Reducción del 87% en tiempo de procesamiento*** *(de 8 horas a 1 hora semanal)* * ***Eliminación de errores de clasificación*** *mediante modelo ML con 88% de precisión* * ***Escalabilidad operativa:*** *Un contador puede atender 3x más clientes* |
| 2. Objetivos | * *Objetivo general y específicos.*   *Automatizar la descarga, procesamiento, clasificación contable y carga de facturas de la SUNAT en CONCAR CB.*   * ***Reducir en un 30%*** *el tiempo de registro manual de facturas.* * ***Disminuir en un 40%*** *los errores humanos en la clasificación contable.* * ***Automatizar el 80%*** *de las facturas procesadas por periodo.* * ***Mejorar la trazabilidad*** *del proceso con reportes y auditoría automática.* * ***Optimizar la conciliación*** *contable reduciendo duplicados en al menos un 90%.* * ***Agilizar la carga a CONCAR CB****, permitiendo integración directa sin reprocesos.* |
| 3. Metodología | Aplicamos un **metodologia ágil** adaptado a las características específicas de un proyecto de IA con componentes de automatización y cloud.  **Sprints 1-2: Extracción automatizada (4 semanas)**   * Desarrollo de bot RPA con Selenium para scraping de SUNAT * Descarga automática de XML y PDF * Almacenamiento en Cloud Storage con metadata * Dashboard de monitoreo básico * **Entregable:** Bot funcional con 3 clientes piloto   **Sprints 3-4: Procesamiento con IA (4 semanas)**   * Recolección y etiquetado de 1,200 facturas * Experimentación con algoritmos ML (Random Forest, XGBoost) * Entrenamiento de modelo: 88% accuracy * Integración con OpenAI para generación de glosas * Transformación a formato Concar CB * **Entregable:** Modelo desplegado en staging   **Sprints 5-6: Registro y cierre (4 semanas)**   * Generación de archivos de carga para Concar CB * Validaciones y logs de auditoría * Testing end-to-end con 10 clientes piloto * Página comercial y backoffice desplegados * **Entregable:** Sistema completo en producción   Esta metodología híbrida era la **única pertinente** porque alinea las prácticas de trabajo con la naturaleza heterogénea del proyecto: ágil para el producto, automatizada para la infraestructura, experimental para la IA. Metodologías tradicionales (cascada) o solo Scrum habrían fracasado ante la complejidad técnica y la necesidad de validación continua con usuarios reales. |
| 4. Desarrollo | **SPRINT 1: Bot de extracción SUNAT (2 semanas)**   * Desarrollo de scraper con Selenium para portal SUNAT * Implementación de autenticación automatizada * Parser de HTML para extraer links de facturas XML/PDF * Descarga paralela con manejo de reintentos * Tests de integración contra portal de pruebas * **Resultado:** 150 facturas descargadas exitosamente   **SPRINT 2: Almacenamiento cloud (2 semanas)**   * Implementación de upload a Cloud Storage * Sistema de nomenclatura estándar de archivos * Metadata: fecha, RUC, tipo, estado * Dashboard web básico mostrando descargas * API REST para consulta de facturas * **Resultado:** Dashboard operativo, 3 clientes piloto onboarded   **SPRINT 3: ML - Preparación de datos (2 semanas)**   * Recolección de 1,200 facturas reales * Etiquetado manual de 800 (training set) * Feature engineering: extracción de campos XML * Análisis exploratorio de datos (distribución de clases) * Versionado de dataset en GCS * **Resultado:** Dataset limpio y balanceado   **SPRINT 4: ML - Entrenamiento y despliegue (2 semanas)**   * Experimentación con Random Forest, XGBoost, Gradient Boosting * *Integración con OpenAI para generación de glosas* * *Despliegue de modelo como servicio REST* * ***Resultado:*** *Modelo en staging clasificando en tiempo real*   ***SPRINT 5: Integración Concar CB (2 semanas)***   * *Análisis de formato de archivo de carga Concar* * *Generación de archivo TXT con estructura requerida* * *Validaciones: RUC válido, montos coherentes, no duplicados* * *Sistema de logs por cada registro procesado* * *Tests de carga en ambiente de pruebas Concar* * ***Resultado:*** *500 registros cargados exitosamente*   ***SPRINT 6: Testing end-to-end y producción (2 semanas)***   * *Testing completo con 10 clientes piloto* * *Corrección de bugs críticos encontrados* * *Optimización de performance (latencia <500ms)* * *Página comercial desplegada (landing page)* * *Backoffice para clientes (consulta de facturas, reportes)* * *Documentación de usuario final* * ***Resultado:*** *MVP en producción, 10 clientes activos*   ***Elementos facilitadores***  ***1. Recursos tecnológicos provistos:***   * ***Licencias RPA gratuitas*** *para desarrollo de bots* * ***Máquinas virtuales GCP*** *proporcionadas por empresa asociada*   ***Acceso a consultores IA/ML*** *para validación de arquitectura del modelo*  ***2. Metodología ágil:***   * *Sprints cortos permitieron detectar problemas tempranamente* * *Daily standups mantuvieron sincronización del equipo* * *Retrospectives generaron mejoras continuas (ej: implementar pair programming en Sprint 3)*   ***3. Clientes piloto comprometidos:***   * *Feedback semanal de alta calidad* * *Disposición a probar versiones beta* * *Datos reales para entrenamiento del modelo*   ***4. Stack tecnológico maduro:***   * *Python con ecosistema robusto (sklearn, pandas, selenium)* * *Docker para portabilidad entre ambientes* * *GitHub Actions con amplia documentación*   ***Dificultades enfrentadas***  ***Dificultad 1: Cambios inesperados en portal SUNAT***   * *En Sprint 1, SUNAT cambió estructura HTML del portal* * *Bot dejó de funcionar durante 2 días* * ***Impacto:*** *Retraso en testing con clientes piloto*   ***Dificultad 2: Precisión inicial del modelo ML bajo 85%***   * *Primer modelo (Sprint 3) alcanzó solo 78% accuracy* * *Clases desbalanceadas (70% gastos, 10% activos, 20% insumos)* * ***Impacto:*** *Necesidad de recolectar más datos y re-entrenar*   ***Dificultad 3: Formato de Concar CB no documentado***   * *Especificación oficial incompleta* * *Tuvimos que hacer ingeniería inversa desde archivos de ejemplo* * ***Impacto:*** *1 semana adicional en Sprint 5*   ***Dificultad 4: Latencia alta en inferencia ML***   * *Primeras versiones tardaban 2-3 segundos por factura* * *Clientes reportaron lentitud inaceptable* * ***Impacto:*** *Experiencia de usuario negativa en staging*   ***Dificultad 5: Coordinación de equipo distribuido***   * *2 miembros en zona horaria diferente* * *Dificultas para pair programming sincrónico* * ***Impacto:*** *Menor colaboración en código complejo* |
| 5. Evidencias | * ***Documentación del proyecto***   *Informe técnico completo (este documento)*  *Diagrama de arquitectura del sistema (flujo completo de datos)*  *Product Backlog con User Stories priorizadas*  *Definition of Done acordada por el equipo*   * ***2. Código y repositorio***   *Repositorio Git con estructura del proyecto*  *Ejemplos de código clave (bot de scraping, modelo ML, API)*  *Pull Requests con code reviews documentados*  *README con instrucciones de instalación y uso*   * ***3. Pipeline CI/CD***   *Archivo .github/workflows con configuración del pipeline*  *Capturas de pantalla de GitHub Actions mostrando builds exitosos*  *Logs de despliegues automáticos a los 3 ambientes*   * ***4. Modelo de Machine Learning***   *Notebook Jupyter con exploración de datos y entrenamiento*  *Métricas finales: Accuracy 88%, Precision 0.86, Recall 0.85, F1 0.85*  *Confusion matrix mostrando performance por clase*  *Gráficos de feature importance*   * ***5. Resultados de validación con clientes piloto***   *Tabla resumen: 10 clientes, tiempo ahorrado (87%), NPS (8.3/10)*  *Testimonios anónimos de 2-3 clientes sobre el impacto*  *Comparativa antes/después en métricas clave*   * ***6. Capturas de pantalla del sistema funcionando***   *Dashboard mostrando facturas descargadas y clasificadas*  *Ejemplo de archivo generado para Concar CB*  *Logs del sistema procesando facturas en tiempo real*   * ***7. Video demo (3-5 minutos)***   *Flujo completo end-to-end: descarga → clasificación → registro*  *Narración explicando cada etapa*  *Sistema funcionando en ambiente de producción real*   * ***8. Documentación de Sprints***   *Burndown charts de 2-3 sprints representativos*  *Actas de retrospectives con acciones tomadas*  *Evolución de velocity del equipo (28 → 42 Story Points)*   * ***9. Presentación ejecutiva (PPT/PDF)***   *Problema, solución, arquitectura, resultados*  *Métricas de impacto y proyección financiera*  *10-15 slides para presentación oral*  ***10. Evidencia de infraestructura***  *Captura de Google Cloud Console mostrando VMs activas*  *Configuración de Cloud Storage con buckets organizados*  *Dashboard de monitoreo básico (uptime, logs)* |
| 6. Intereses y proyecciones profesionales | * *Al inicio de la asignatura, mi interés era genérico: "quiero desarrollar software". Este proyecto me obligó a enfrentarme a la realidad completa de crear un producto tecnológico, y eso cambió todo.* * *Descubrí que lo que realmente me apasiona no es solo escribir código elegante, sino* ***ver cómo ese código resuelve problemas reales de personas reales****. Cuando el primer cliente piloto nos dijo "esto me ahorró 6 horas esta semana", sentí una satisfacción que ningún examen aprobado me había dado antes.* * *También descubrí que disfruto enormemente el desafío de* ***poner IA en producción****. Entrenar un modelo que funciona en un notebook es relativamente fácil; hacer que ese modelo clasifique 400 facturas diarias sin fallar, que se mantenga preciso en el tiempo, que responda en menos de 500ms... eso es ingeniería real y me fascina.* * *Continuar con este proyecto y* ***convertirlo en una startup real****. Tenemos todo lo necesario: producto validado, clientes pagando, mercado de 50,000 empresas potenciales, y un equipo técnico sólido. El próximo paso es postular a una incubadora (Start-Up Chile, UTEC Ventures) para conseguir capital semilla y escalar a 100 clientes en 6 meses.* * *En este escenario, asumiría el rol de* ***CTO/Co-founder****, liderando el desarrollo técnico mientras aprendo sobre el lado comercial del negocio.* |