**Actividad: Propuesta de desarrollo y despliegue de un proyecto de inteligencia artificial**

*Título del proyecto*: **Aplicación de asistente personalizado de asesoramiento nutricional utilizando aprendizaje automático diseñado para la población de El Salvador.**

**Introducción y requisitos del proyecto.**

Existen muchas propuestas en el mercado de aplicaciones orientadas a fitness, salud y nutrición. Sin embargo, la mayoría de ellas no involucran en sus dietas, alimentos de consumo local en cada país, por lo que generalmente los usuarios terminan desmotivados al no poder conseguir los requisitos de una dieta determinada y abandonan la aplicación.

El objetivo de este proyecto es **crear una aplicación personalizada para la población local** que permita recomendar una mejor dieta para alcanzar los propósitos que cada usuario desee, como bajar de peso, verse mejor o aumentar masa muscular.

Para esto, la aplicación deberá analizar la calidad nutricional de los alimentos que consume el usuario, el índice de masa corporal (IMC), edad, sexo, y en base a eso *generar recomendaciones personalizadas, es decir,* una dieta balanceada, que sería equivalente a la recomendación de un nutricionista.

Se pretende obtener dos rasgos diferenciadores de la aplicación:

1. La aplicación debe dar prioridad a alimentos de producción local, la mayoría de las dietas de la competencia no poseen bases de datos con alimentación específica del país por lo que algunas recomendaciones de productos no se encuentran disponibles en el mercado, generando insatisfacción de los usuarios.
2. La aplicación debe ser capaz de usar procesamiento del lenguaje natural. Cuando el usuario ingrese su dieta regular, este debe de poder hacerlo en el lenguaje cotidiano y será la aplicación la encargada de procesar la información.

**Implementación.**

Para llevar a cabo el proyecto se necesitará construir primeramente una base de datos relacional de los alimentos consumidos localmente, en la base de datos debe poderse categorizar la producción local, productos importados y también toda la información nutricional de los mismos.

Además, debe diseñarse una dieta alimenticia basada en los macronutrientes necesarios para el cuerpo humano, la cual será caracterizada a partir de la estatura, sexo, edad, y dieta actual seguida por el usuario.

En tercer lugar, se necesitarán datos de entrenamiento para poder crear un modelo de recomendaciones utilizando *machine* *learning (ML)*. Esto requiere un estudio con un muestreo de datos que tenga representación significativa en la población objetivo de la aplicación, con edades de 16 a 60 años. Para crear el modelo nutricional se requerirá de la asesoría de un nutricionista.

Una vez se haya comprobado la validación del modelo se procederá a realizar pruebas con nuevos datos, no conocidos en la etapa de entrenamiento, el objetivo es hacer una búsqueda de grilla de hiperparámetros con validación cruzada (*GridSearchCV*) hasta llegar a métricas que garanticen la calidad de las dietas recomendadas por la aplicación.

Con un modelo funcional, el siguiente paso será la implementación del procesamiento de lenguaje natural para el ingreso de datos por parte del usuario. El objetivo es que sea la aplicación capaz de tomar datos de entrada del usuario de la forma más simplificada posible.

La etapa final consistirá en la implementación de la aplicación en la web.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nombre de la tarea | Fecha de inicio | Fecha de finalización | Asignado | Estado | SEMANA 1 | SEMANA 2 | SEMANA 3 | SEMANA 4 | SEMANA 5 | SEMANA 6 | SEMANA 7 | SEMANA 8 | SEMANA 9 | SEMANA 10 | SEMANA 11 | SEMANA 12 | SEMANA 13 | SEMANA 14 | SEMANA 15 | SEMANA 16 | SEMANA 17 | SEMANA 18 | SEMANA 19 | SEMANA 20 |
| **PROYECTO: WEBAPP NUTRICION** | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Investigación de campo | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Base de datos | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Modelo nutricional | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Entrenamiento de modelos de ML | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Evaluación y mejoramiento de hiperparámetros. | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Procesamiento del lenguaje natural | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Desarrollo de la app vía PaaS | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Pruebas | ####### | ####### |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Tabla 1**

*Diagrama de Gantt de las principales etapas del proyecto.*

*Nota*. Esta tabla muestra la línea de tiempo a seguir basado en las actividades descritas en la sección de implementación. Elaboración propia.

**Recursos necesarios.**

A continuación, se presentan los recursos necesarios en el desarrollo del proyecto, la lista está sujeta a variaciones según la línea de tiempo de ejecución del proyecto, avance o surjan imprevistos:

1. *Recurso humano*: Será fundamental dividirlo en equipos y asignar tareas específicas a cada uno para lograr resultados eficientes en el menor tiempo posible.
2. *Software*: Será fundamental usar PaaS (Platform as a Service) como la herramienta para el diseño de todo el proyecto por su facilidad de dotar toda la infraestructura para el funcionamiento de la aplicación, por tal razón el *Data Architect* realizará las valoraciones de costo-beneficio de implementación en las principales plataformas (AWS, Google Cloud, Azure) y seleccionará la más adecuada al proyecto.
3. *Investigación de campo*: Será importante para la implementación de la base de datos de alimentos personalizada para el país. Se tomará como punto de partida las encuestas de alimentos de la canasta básica realizadas por la Oficina Nacional de Estadísticas y Censos (ONEC).
4. *Instalaciones*: El trabajo será remoto con sesiones diarias basadas en la metodología scrum para informar los avances, las nuevas actualizaciones del proyecto y los próximos objetivos.
5. *Financiamiento*: Inversión propia de la empresa con un capital inicial de $10,000 sujeto a los análisis sobre el riesgo/beneficio del proyecto.

**Técnicas de ML a utilizar.**

Para la generación de un modelo que permita generar una dieta personalizada se realizarán pruebas para encontrar el modelo que mejor se ajuste a los datos sin causar un *overfit*. Uno de los algoritmos de clasificación más comunes es *K-Means*, que puede ser utilizado para el agrupamiento de usuarios y alimentos. El propósito es clasificar la alimentación vía macronutrientes y calorías ingeridas. Otro algoritmo por considerar es *Random Forest Classifiers*, de igual manera se buscará poder generar un sistema de recomendaciones lo más cercano al que podría ser entregado por un nutricionista.

Se utilizará, además, *procesamiento de lenguaje natural* (*NLP*) para la extracción de todas las features relevantes que pueda ingresar el usuario, tales como, alimentos, cantidades, bebidas, etc.

***Metodología de desarrollo.***

Scrum, es un marco de trabajo utilizado generalmente en desarrollos de proyectos de *software*. Aunque no es una metodología, su flexibilidad de trabajar en pequeños *sprints* o incrementos, lo hace ideal en un proyecto o producto complejo como el proyecto presentado. Scrum también está diseñado para generar un MVP (producto de mínimo esfuerzo) en el menor tiempo posible.

Por las razones anteriormente descritas, será Scrum el principal marco a seguir, pero también se usarán los diagramas de Gantt para poder dirigir la línea de tiempo del proyecto.

Se crearán equipos de trabajo bajo el marco Scrum, donde cada *Scrum* *Team* siga una meta determinada del *Product* *Backlog*, o requerimientos del producto, los equipos serán homogéneos y entenderán las metas a seguir sin sentir que no pertenecen a un determinado equipo de trabajo.

Bajo el marco Scrum también se definirán los requisitos del producto, los incrementos que se espera realizar en cada *sprint* y adecuar el timeline del proyecto en base a los resultados y el *feedback* del cliente.

De las características anteriores, podemos enumerar los roles necesarios a considerar para la puesta en marcha del proyecto:

1. Un *Scrum Máster* que presentará todo el marco a seguir a todos los involucrados.
2. Un *Product Owner* encargado de gestionar el producto final a través de la actualización constante del backlog o requisitos del sistema.
3. *Scrum* *Teams* de entre 5 a 7 personas máximo, encargadas de desarrollar el producto o algunos de los requerimientos del proyecto establecidos en el cronograma a seguir.

**Recurso humano.**

El equipo debe ser multidisciplinario formado por:

* *Científicos de datos* para realizar las tareas de recolección de datos, limpieza, filtrado de características y entrenamiento y validación de los modelos de *Machine Learning*.
* Nutricionista para brindar asesoría en el desarrollo del producto por parte de un experto en el tema. La validación de los modelos anteriores también debe ser parte de esta asesoría.
* *Programadores Full-Stack* encargados de la aplicación, desarrollo de la interfaz gráfica y pruebas del software en desarrollo y producción.
* *Cloud Architect*, encargado de seleccionar la infraestructura más adecuada y de la implementación de la aplicación vía PaaS.

**Despliegue (*deployment*) del proyecto.**

Para realizar el *deployment* del proyecto se realizará su implementación en la nube usando PaaS en cualquiera de las plataformas más reconocidas. Una plataforma por considerar es Heroku por ser ampliamente conocido como una herramienta de despliegue de aplicaciones web en la nube.

En cuanto a herramientas de software a utilizar en desarrollo, se realizará la aplicación web utilizando *Django*, uno de los frameworks más robustos para el desarrollo web en Python. Los modelos de ML utilizados serán implementados de las librerías de *sklearn* para utilizar el algoritmo de Random Forests, *spacy* para depuración y pruebas de lenguaje natural y *nltk* para entrenamiento del algoritmo de NLP (Natural Language Processing).

**Figura 1**

*Diagrama de deployment del proyecto.*

Diagrama

Descripción generada automáticamente

*Nota*. El modelo tiene como propósito garantizar la comunicación entre módulos para una mejor operatividad. Elaboración propia.

De la figura 1, se debe prestar importancia a los siguientes elementos:

* La autenticación de usuarios se realizará vía *Open Authentication* (OAuth) en los servidores propios de la empresa antes de enviar la información restante a la nube, esto con el fin de garantizar la integridad de cada cuenta y evitar el acceso no autorizado.
* La infraestructura principal de la aplicación se almacenará en la nube, tanto los servidores de almacenamiento de la aplicación y la base de datos relacional utilizando MySQL se tendrán en la nube.
* En las rutinas de *testing* (pruebas) se deberá utilizar Git como herramienta de control de versiones que permita escalar la aplicación a medida su desarrollo vaya completándose, será fundamental hacer pruebas de la UI (interfaz de usuario) y la API (la comunicación con los servidores remotos) y llevar un registro de los resultados de las pruebas.

**Seguridad y confidencialidad de la información.**

Toda información de los usuarios que pueda servir para la identificación de una persona será considerada como confidencial, y será tratada de acuerdo con la normativa vigente del país.

La seguridad de los datos es también un aspecto fundamental en el desarrollo del proyecto, por esa razón será necesario implementar mecanismos para controlar el acceso no deseado, como la autenticación descrita anteriormente.

La seguridad de toda la infraestructura en la cual se aloja el proyecto será provista por el proveedor de servicios en la nube bajo la modalidad PaaS.

En ningún caso será necesario requerir más información de la estrictamente útil para generar los modelos de nutrición, de esta manera se evitará el reconocimiento de terceros de un usuario garantizando la anonimización de los datos.

**Bibliografía**.

* Schwaber K., Sutherland J. (2020). *La Guía Scrum: La Guía Definitiva de Scrum: Las Reglas del Juego*. https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-Spanish-European.pdf
* UNIR. (2022c, diciembre 14). *¿Qué es el despliegue de aplicaciones web?* https://www.unir.net/ingenieria/revista/despliegue-aplicaciones-web/