**UNIVERSIDAD MAYOR, REAL Y PONTIFICIA DE**

**SAN FRANCISCO XAVIER DE CHUQUISACA**

**FACULTAD DE TECNOLOGIA**



**TÍTULO DEL PROYECTO** Interprete del lenguaje de señas

**ÁREA** Inteligencia Artificial

**CARRERA o INSTITUCIÓN** Ing. Ciencias de la Computacion

**NOMBRE** Celedonio Perka Casillas

**TUTOR** Ing. Carlos Walter Pacheco Lora

**SUCRE – BOLIVIA**

**2022**

Indice

1. Indice 4

2. Resumen 4

3. Antecedentes 4

4. Problematica 4

5. Abordaje 4

6. Objetivo General 4

7. Objetivos Especificos 5

8. Fundamentos Teoricos Conciderados en el trabajo 5

9. Metodologia 5

10. Resultados 6

**Interprete del Lenguaje de Señas**

## Resumen

Un software inteligente que interpreta las señas una persona sordomudo (discapacitado) que habla en lenguaje de señas, para personas que no conocen el significado de las señas y mostrar el significado en texto (Se pretende interpretar las señas no mas de 10 clases debido a que se cuenta con recursos limitados y por falta del personal para la creacion del dataset).

## Antecedentes

Las personas con la discapacidad de sordomudo a menudo se sienten excluidos en la sociedad por no poder comunicarse facilmente.

La mayoria de las personas no hablamos o entendemos el lenguaje de señas que es precisamente la causa de este problema de exclucion social.

## Situación Problematica

En bolivia hay aproximadamente 90.000 personas sordas hasta el 2021 y muchas de ellas no tienen acceso a la educacion.

La exclusion social es uno de los problemas principales para estas personas,

al no poder comunicarse sufren depresion.

No existe software comercial para este proposito.

Los softwares propuestos estan en desarrollo por empresas tecnologicas.

## Problema Principal

Las personas con discapacidad auditiva sufren aislamiento social, sin acceso a la eduacion, desempleo o que realizan mayor fuerza de trabajo en su labor por tratar de comunicarse con los demas.

1. **Objetivo General**

Desarrollar un modelo inteligente que permita la comprension a las personas con discapacidad auditiva que hablen en lenguaje de señas, hacia las personas que no comprendemos el lenguaje de señas, utilizando un modelo inteligente, integrar a la sociedad a las personas sordomudas atra vez de la posibilididad de comunicacion utilizando un modelo inteligente para el interprete de lenguaje de señas a texto.

## Objetivos Específicos

* Analizar las tecnicas utilizadas actualmente para este ambito de interprete de señas y adaptarlo al lenguaje de señas boliviano.
* Aplicar arquitecturas para la clasificacion de datos numericos como el perceptron multicapa o arquitecturas livianos debido a que el modelo se ejecutaran en tiempo real.
* Desarrollar un modelo de Deep Learning que aprenda los gestos y que clasifique a texto el gesto como puntos clave de entrada (video de entrada)
* Desplegar en un servidor el modelo entrendo cuando se alcanzo una precision mas del 90% en interpretar señas y en la generacion de texto con las palabras mas basicas.
* Validar el modelo en un entorno real usando la web cam o la camara del celular para capturar los gestos a una distancia no mas de 1,50cm.

## Fundamentos Teóricos Considerados en el trabajo

1. **Ambito de la inteligencia artificial, tecnicas, algoritmos, entre otros**

La lengua de señas es la lengua natural de las personas sordas.

Una lengua que como cualquier otra, posee y cumple todas las leyes lingüísticas.

La comunicacion se realiza principalmente con expresiones gestos de las manos, expresiones faciales y gestos corporales.

Para señas de una sola mano se utiliza la mano dominante y para señas mas complejos las dos manos.

1. **Ambito al que se aplico la inteligencia artificial**

No existe un softare comercial como interprete del lenguaje de señas.

Las propuestas desarrolladas estan limitadas al reconocimiento de gestos estaticos y los modelos que interpretan gestos dinamicos estan limitadas a muy pocas clases no mas de 10 clases o palabras.

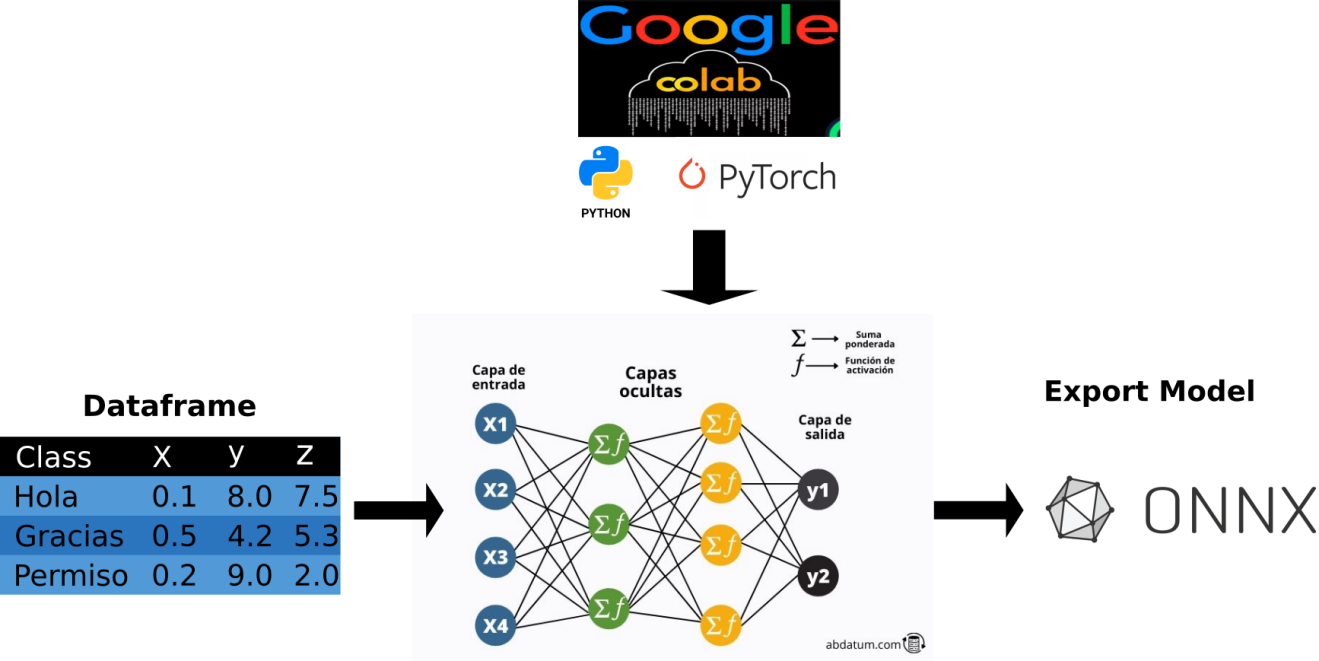
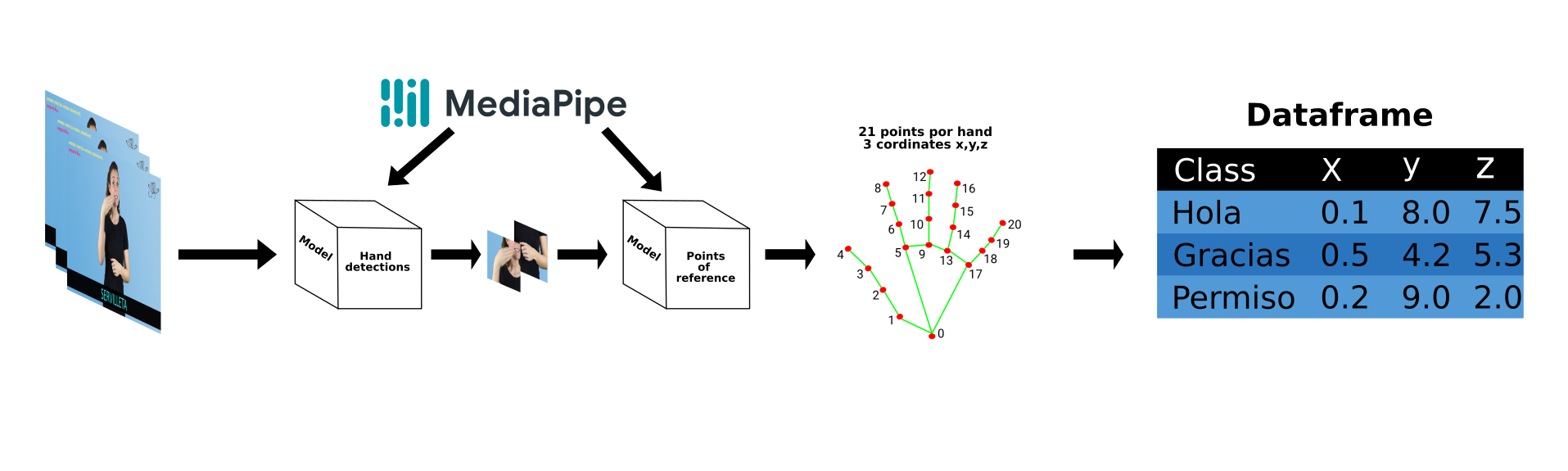
La solucion a este problemas es muy desafiante y requiere muchos recursos.

Showleap empresa española tiene hasta ahora uno de los mejores modelos para la traduccion, que consta de tres deredes neuronales uno procesa video, las egunda capta los signos e interpreta y la tercera crea la frase para con sentido.

## Componente IA

1. **Descripcion y esquema del modelo o modelos**

El modelo para capturar puntos clave es propiedad de Google, MediPipe Hands, el cual devuelve puntos con lo cual se entrena la nueva red que clasifica un grupo de puntos (frames)



1. **Descripcion de adquisicion de datos y preprocesamiento**

Se grabo videos realizando los gestos y el autor del video y gesto es mi persona. (Celedonio Perka Casillas)

Para grabar los videos se uso un celular Samsung Galaxy A50 con 4GB de RAM y la camara de 32MP.

El dataset de los videos grabados se procesa pasando por un editor de video, recortar en videos cortos por cada gesto y reducir los frames de los videos a 15fps para reducir la sobrecarga del modelo.

Cada uno de los videos se guarda en un directorio especifico que se seran leidos para la extraccion de puntos clave de los frames con un modelo entrenado de Google MediaPipe Hands que devuelve coordenadas que se guardaran en un dataframe.

Los gestos con frames (puntos clave) cortos son completados con valores cero a la longitud del gesto mas largo y posteriormente guardar en un archivo csv.

1. **Técnicas y metricas de entranamiento**

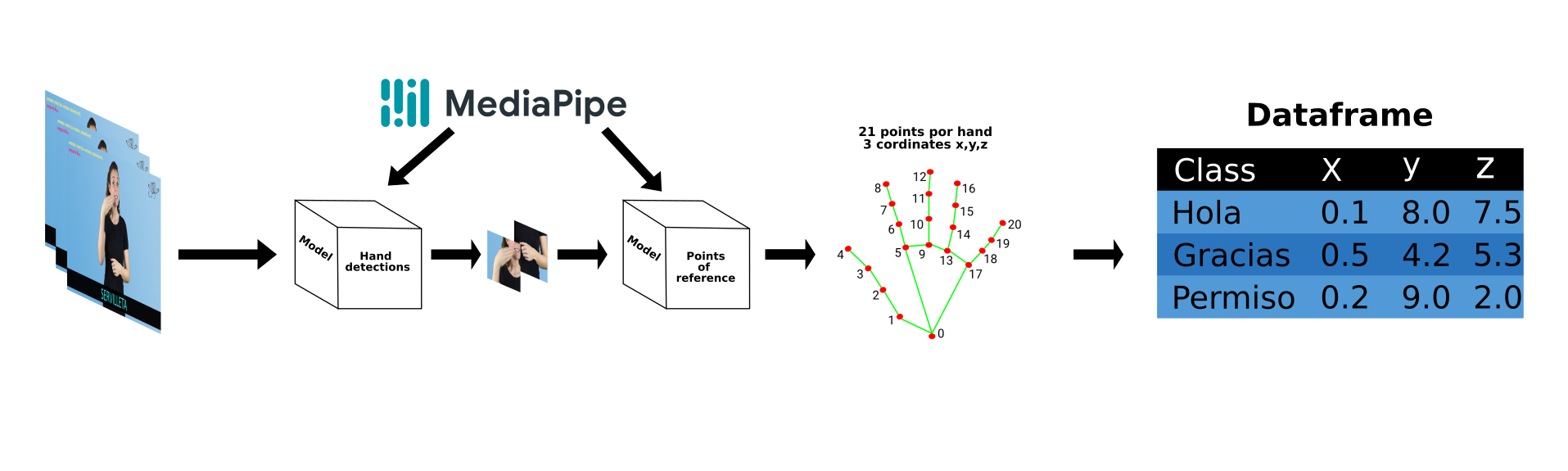
Las tecnicas utilizadas para entrenar el modelo, separar datos para entrenamiento evaluacion y test, entrenar con diferentes hiperparametros y aplicar los mejores hiperparametros que logren la presicion mas alta, que a su vez tenga la latencia mas corta posible debido a que debe ejecutarse el modelo en tiempo real.

**Precision:** de los 248 videos (gestos) en 80% fue para entrenar el modelo y el 20% para evaluar con lo cual obtuvimos un 95% de precision con los datos de evaluacion.

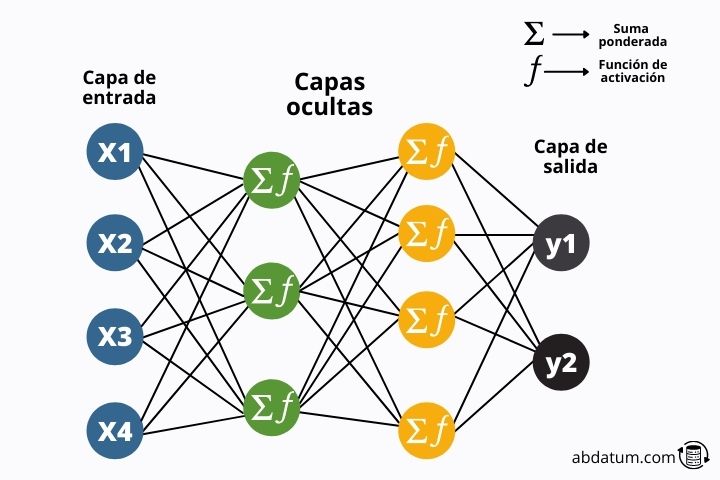
## Descripcion de trabajo realizado

1. **Esquema y descripcion de arquitecturas, frameworks**

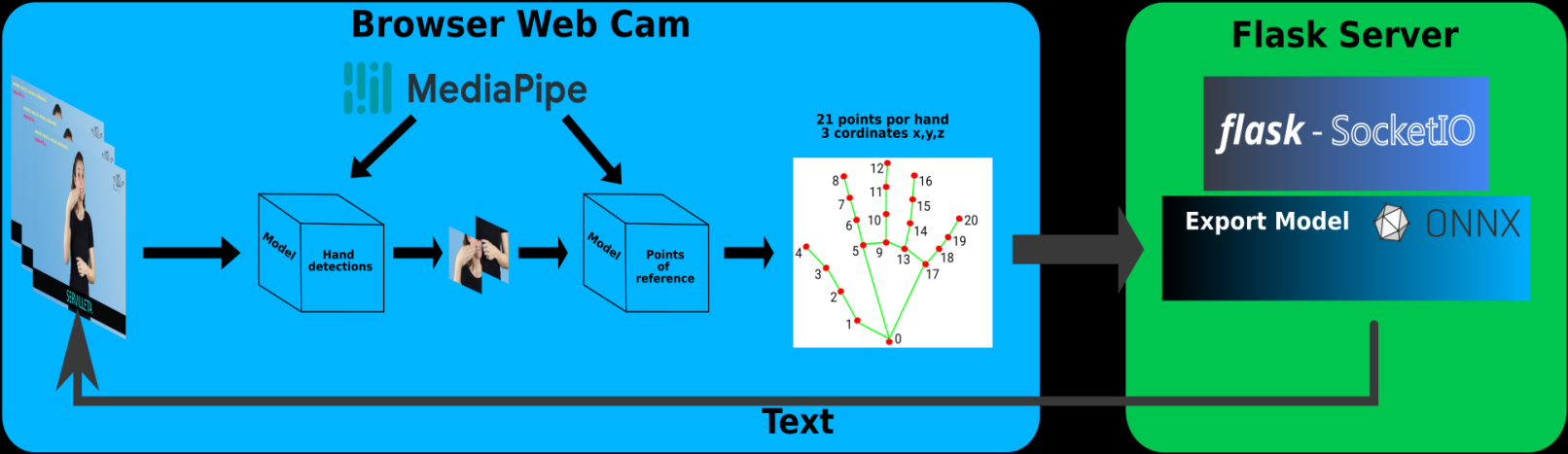
Red preentrenada MediaPipe Hands de Google que caputara puntos clave de una imagen que contiene palmas de la mano



Se uso un perceptron multicapa funcines de activacion RELU en capas intermedias y la funcion softmax en la capa de salida y la probabilidad mas alta como resultado.

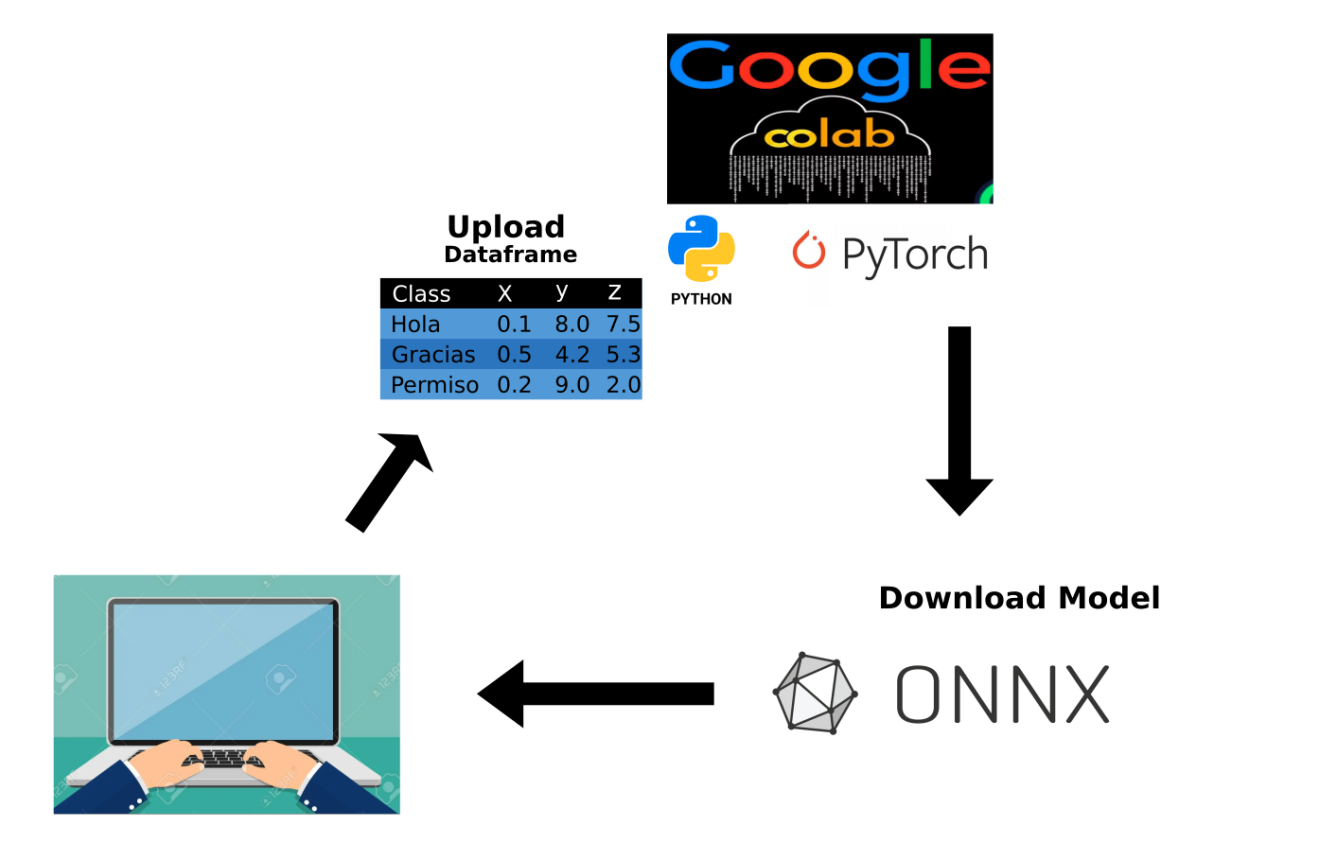


Aplicacion desplegado en un servidor con Flask, Python y Socket.IO para la tranferencia de datos en tiempo real entre el cliente y el servidor.



1. **Esquema y descripcion componentes de hardware empleados**

Los componentes empleados fueron una portatil personal para el preprocesamiento de los datos y la nube de Google para entrenar e modelo Colab.



1. **Herramientas utilizadas (Hardware y Software)**

El preprocesamiento de datos en una maquina portatil normal de uso personal.

Con 16GB de RAM sin tarjeta grafica dedicada y un procesador Core i7 de 8va geberacion

Para el entrenamiento del modelo se principalmente Google Colab

debido a que nos brinda una GPU.

Visual Studio Code como ID para revision y desarrollo de preprosesamiento de datos y el modelo mismo.

1. **Cronograma**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nro | Actividades | Abril | | | | Mayo | | | | | Junio | | | | | Julio | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | | 2 | 3 | 4 | 1 | | 2 | 3 | 4 |
| Paso 1 | Preparar el dataset |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |
| Paso 2 | Preprocesamiento del dataset |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |
| Paso 3 | Desarrollar el modelo |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |
| Paso 4 | Entrenar y evaluar el modelo |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |
| Paso 5 | Poner en produccion el software |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |

1. **Resultados obtenidos**

De los 303 videos de gestos el 80% se uso para el entrenamiento del modelo y el 20% restante para evaluar el modelo.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Modelos | Capas ocultas | Nro de perceptron promedio | Learning rate | Precision |
| MLP | 2 | 1500 | 0.08 | 95 |

## Conclusiones

Se encontro un problema al momento de leer video por la web cam se requiere frames constante debido a que el modelo fue entrenado con videos de 15 fps.

Los frames fluctuan dependiendo de los recursos del dispositivo en la cual se ejecuta.

1. **Referencias Bibliograficas**

Universidad Catolica Boliviana 2021

<https://sobretodopersonas.ucb.edu.bo/2021/07/11/sistema-de-reconocimiento-de-la-lengua-de-senas-boliviana/>

https://medium.com/saturdays-ai/detector-gestual-del-lenguaje-de-signos-espa%C3%B1ol-a74d15b6bd9e