Hola YouTube, en este video te mostraré cómo puedes convertir rápidame
nte cualquier audio en texto usando el paquete de código abierto gratuito
en Python llamado Whisper.Voy a mostrar que lo instalé, mostrar un ejemp
lo de cómo lo ejecuté y compararlo con una biblioteca existente.Entonces,
comenzando, probablemente querrá ir al repositorio Whisper Get Hub que
estamos viendo aquí y dan instrucciones sobre cómo puede instalarlo.Ahor
a, una cosa a tener en cuenta, cuando PIP instale solo el nombre, no va a in
stalar la versión correcta.Queremos instalar desde este repositorio Git.Así
que solo tome este comando de instalación de PIP y ejecute en su entorno
que esté ejecutando Python.Y también mencionaron aquí que necesita inst
alado FFM PEG.Hay algunas instrucciones para hacerlo, pero ya lo tuve inst
alado en mi computadora.Ahora que tengo una instalación de susurro, hag
amos un poco de audio en el que pueda probar esto.Así que voy a decir alg
unos modismos.Los modelos suelen ser difíciles de entender para los mod
elos.A pesar de que esto es solo discurso a texto.Esto será un poco divertid
o.Me encantaría estar en la nube 9 como un pony único que no dolería una
mosca.Sería como un pez fuera del agua y tan en forma como un violín par
a estar bajo el clima.Guardemos esto.Guardemos como una ola.Tienen inst
rucciones sobre cómo podríamos ejecutar esto simplemente directamente
desde la línea de comando una vez que está instalada.Te mostraré cómo u
sar la API de Python, que muestran aquí.Entonces es realmente simple.Sol
o importamos susurros.Entonces vamos a crear nuestro modelo, que vamo
s a cargar.modelo que se llama base.Y luego solo usando este objeto mode
lo, ejecutamos Transcribe en nuestro archivo de audio.Entonces lo llamé m
odismos.Usemos la versión de onda.Queremos que esto devuelva el result
ado.Ahora, me di cuenta de que cuando corrí esto antes, recibo este error
debido a la mitad tensor de Kuda y el tensor flotante.Pude resolver esto.En
tonces eso es algo a tener en cuenta.Si no funciona para usted, es posible
que deba establecer el punto flotante 16 para caer.Y puede ver después de
que se ejecute aquí, detectó el idioma ya como inglés y luego este objeto d
e resultado tiene algunos métodos diferentes en ellos, pero lo que querem
os dentro de esto es solo el texto y podríamos ver que parece que se ve.Al i
gual que el resultado, es bueno, me encantaría estar en la nube nueve com
o un pony de un truco que no dolería una mosca, sería como un pez fuera
del agua y esto se arruinó un poco de este pez de agua enTan en forma co
mo un violín y tal vez no lo dije con claridad, otra cosa es saber es cuando e
jecute esto por primera vez, tendrá que descargar el modelo base.Por lo ta
nto, es posible que vea una barra de progreso y tendrá que descargar ese
modelo.Y dice que cuando ejecuta esta transcripción, en realidad está tom
ando 30 segundos de su archivo de audio y ejecutando predicciones en él.
Ahora también hay otro enfoque que puede adoptar, que es un enfoque d
e nivel inferior, donde realmente crea el modelo y luego crea el objeto de a

udio y el patrón recorta esto.Entonces, solo asegúrese de que este fragme
nto de audio sea de solo 30 segundos.segundos o lo dará una palmada con
30 segundos, ya que esa es la longitud que el modelo espera tener como e
ntrada.Luego está haciendo un espectrograma del mouse log.Está detecta
ndo el lenguaje y podemos decodificar aquí y proporcionar muchas más op
ciones si quisiéramos.Si ejecuto esta celda, nuevamente obtenga este error
, que ahora puedo configurar en las opciones de decodificación, FP16 es ig
ual a fallas.Y en realidad esta vez parece que todo tiene todo correcto.Sería
como un pez fuera del agua.y está en forma como un violín.Así que eso es t
odo para susurros.Solo quiero compararlo con un tipo de modelo existente
.Y una biblioteca popular para hacer esta es la biblioteca de reconocimient
o de voz.La forma en que ejecutamos la biblioteca de reconocimiento de vo
z es que la importamos y luego creamos este objeto reconocedor, con el q
ue luego podemos cargar nuestro archivo de audio.Después de eso, puede
tomar el objeto reconocedor y hay algunos métodos de reconocimiento dif
erentes para eso.Y vamos a usar el reconocimiento de Google y veamos cu
ál es el resultado.Por lo tanto, parece que no agregó ninguna puntuación,
y la nube nueve es diferente.Me encantaría estar en la nube nueve como u
n pony único que no dolería una mosca.Pero lo único a tener en cuenta es
que esto realmente está usando la API de reconocimiento de voz de Googl
e.La biblioteca Whisper, en realidad se descarga el modelo y es suyo para u
sar.También le recomiendo que eche un vistazo al documento Whisper, qu
e se lanzó con este código.También entran en detalles sobre cómo se entre
nó el modelo y la arquitectura que se usa.Whisper trabaja en un montón d
e idiomas diferentes.El rendimiento que dicen varía según el idioma.Así qu
e puedes ir aquí en el repositorio de GitHub, donde tienen una trama que
muestra qué idiomas realmente funcionan mejor para los bares aquí.Más
pequeño es mejor y más grande significa que funciona peor.Así que sigue
siendo bastante impresionante la cantidad de idiomas en los que este mod
elo funciona.