



Title: Recommendation of CFDI-affected accounts based on machine learning

Authors: MONTECILLO-PUENTE, Francisco Javier, PERÉZ-MONCADA, Santiago
 and VÁZQUEZ-CHÁVEZ, Luis Alejandro

Editorial label RINOE: 607-8695
BECORFAN Control Number: 2023-03
BECORFAN Classification (2023): 111213-0301

Pages: 16

RNA: 03-2010-032610115700-14

RINOE - Mexico

Park Pedregal Business. 3580-
 Adolfo Ruiz Cortines Boulevard –
 CP.01900. San Jerónimo Aculco-
 Álvaro Obregón, Mexico City
 Skype: MARVID-México S.C.
 Phone: +52 1 55 6159 2296
 E-mail: contact@marvid.org
 Facebook: [MARVID-México S. C.](https://www.facebook.com/MARVID-México-S-C)
 Twitter: [@Marvid_México](https://twitter.com/Marvid_México)

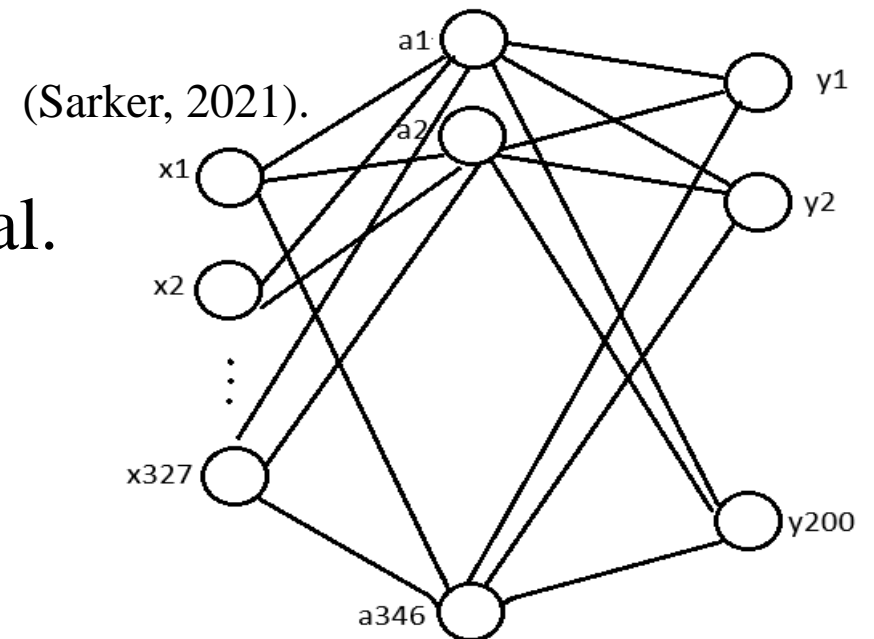
www.rinoe.org

Holdings

Mexico	Peru
Bolivia	Taiwan
Cameroon	Western
Spain	Sahara

Introducción: Aprendizaje automático basado en redes neuronales. RNN

- Generar datos y su etiquetado.
- Separar en datos de prueba y validación.
- Pre procesar datos.
- Definir el modelo de red neuronal.
- Entrenar la red.
- Validar la red.



Introducción: SAT facturación electrónica CFDI.

This XML file does not appear to have any style information associated with it. The document tree is shown below.

SAT, 2023

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<cfdi:Comprobante xmlns:cfdi="http://www.sat.gob.mx/cfd/4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" Exportacion="01" LugarExpedicion="58640" MetodoPago="PUE"
TipoDeComprobante="I" Total="496.79" Moneda="MXN" SubTotal="396.55"
Certificado="MIIFdjCCA16gAwIBAgIUzAwMDEwMDAwMDA0MDAwMDIzMjUwDQYJKoZIhvcNAQELBQAwggErMQ8wDQYDVQQDDAZBQyBQVQVxLjAsBgNVBAoMJVNFU1ZJQ00PIERFIEFETU10SVNUUkFDSU90IFRSSUJVVFESSUEXGjA'
NoCertificado="30001000000400002325" FormaPago="01"
Sello="FRBm2S3Ls2E8dKPKSqiXm0ZpnuCbj2Iw56KqeImUKpMYp+YueA7PUBf0ZyhCq7135VBjDpZfYknfLDPVQjuuawZMMvfSsuNLOYceX8tr1PPDeow/gJmkkj96DdnkmYFKDSh+guTji4C9UU45d7Brmqas7DLMPiXNwYI/iIA6s'
Fecha="2022-09-29T10:07:04" Folio="1657" Serie="A" Version="4.0" xsi:schemaLocation="http://www.sat.gob.mx/cfd/4 http://www.sat.gob.mx/sitio_internet/cfd/4/cfdv40.xsd">
  <cfdi:Emisor RegimenFiscal="612" Nombre="XAIME WEIR ROJO" Rfc="WERX631016S30"/>
  <cfdi:Receptor DomicilioFiscalReceptor="58640" RegimenFiscalReceptor="616" UsoCFDI="S01" Nombre="LIONDEV" Rfc="XAXX010101000"/>
  <cfdi:Conceptos>
    <cfdi:Concepto ObjetoImp="02" Importe="396.55" ValorUnitario="19.83" Descripcion="Gasolina Premium" Unidad="Punto de conferencia" ClaveUnidad="Z6" Cantidad="20.00"
    NoIdentificacion="1234" ClaveProdServ="15101515">
      <cfdi:Impuestos>
        <cfdi:Traslados>
          <cfdi:Traslado Importe="68.52" TasaOCuota="0.160000" TipoFactor="Tasa" Impuesto="002" Base="428.28"/>
          <cfdi:Traslado Importe="31.72" TasaOCuota="0.080000" TipoFactor="Tasa" Impuesto="003" Base="396.55"/>
        </cfdi:Traslados>
      </cfdi:Impuestos>
    </cfdi:Concepto>
  </cfdi:Conceptos>
  <cfdi:Impuestos TotalImpuestosTrasladados="100.24">
    <cfdi:Traslados>
      <cfdi:Traslado Base="428.28" Importe="68.52" TasaOCuota="0.160000" TipoFactor="Tasa" Impuesto="002"/>
      <cfdi:Traslado Base="396.55" Importe="31.72" TasaOCuota="0.080000" TipoFactor="Tasa" Impuesto="003"/>
    </cfdi:Traslados>
  </cfdi:Impuestos>
  <cfdi:Complemento>
    <tfd:TimbreFiscalDigital xmlns:tfd="http://www.sat.gob.mx/TimbreFiscalDigital" FechaTimbrado="2022-09-29T10:07:04" NoCertificadoSAT="30001000000400002495"
    RfcProvCertif="SPR190613I52"
    SelloCFD="FRBm2S3Ls2E8dKPKSqiXm0ZpnuCbj2Iw56KqeImUKpMYp+YueA7PUBf0ZyhCq7135VBjDpZfYknfLDPVQjuuawZMMvfSsuNLOYceX8tr1PPDeow/gJmkkj96DdnkmYFKDSh+guTji4C9UU45d7Brmqas7DLMPiXNwYI
    SelloSAT="Gxi/0qGfYTpWvXTziEkHber11clJ6tH4/iVZ1XndRLKHGxjoA4Y8Qf1kQPrKj9oH92BGCv+VHs4ft0TAUKg/Rg13snrmdaFT8JzLHU5JLjhZ18aTM8t0utWCGAmaM7GbLfu0W5+jfJDXk+jjFQNEk1HODDVG1HBXN:
    UUID="4a4d2778-e4d1-4d38-88c1-dc4f32b766bb" Version="1.1" xsi:schemaLocation="http://www.sat.gob.mx/TimbreFiscalDigital
    http://www.sat.gob.mx/sitio_internet/cfd/TimbreFiscalDigital/TimbreFiscalDigitalv11.xsd"/>
  </cfdi:Complemento>
  <cfdi:Addenda>
    <Propina:Propina xmlns:Propina="https://facture.com.mx/xsd/Propina" Tasa="0.100000" xsi:schemaLocation="https://facture.com.mx/xsd/Propina
    https://facture.com.mx/xsd/Propina.xsd"/>
  </cfdi:Addenda>
</cfdi:Comprobante>
```

Introducción

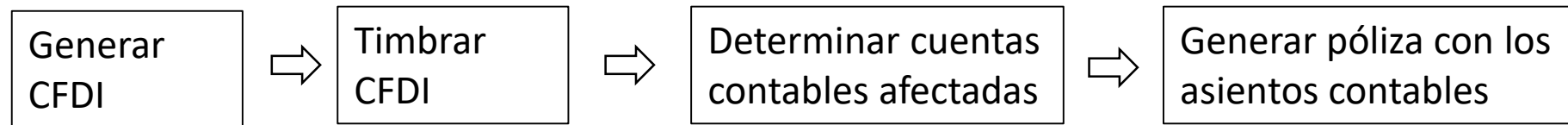
- **FactureApp de LionDev SA de CV**
<https://facture.com.mx/>

The screenshot displays the Facture App interface. On the left, a navigation menu includes sections like 'CONTABILIDAD', 'NÓMINA', 'CARTA PORTE', and 'Administración'. The main dashboard area is divided into several panels: 'Acceso directo facturación', 'Acceso directo catálogos', 'Facturado este mes' (with a table of transactions), 'Recibido este mes' (with a table of receipts), 'Ingresos por sucursal este mes' (with a pie chart), and 'Egresos por emisor este mes'. A right-hand sidebar shows 'Facturas por cobrar vencidas' (with a table of due invoices), 'Su cuenta' (account balance), and 'Noticias' (news). The background of the slide features a blurred image of a person working at a computer.

Facture App
Plataforma digital para facturación, nómina y contabilidad que da solución a tus necesidades contables y fiscales.

Introducción

- Creación de póliza



En este trabajo una red neural es entrenada para definir un sistema de recomendación de pólizas en base a las posibles cuentas contables que pueden modificar los CFDI's.

Metodología: Conceptos de pólizas

- Datos que debe contener una póliza
 - a) Información sobre el tipo de póliza,
 - b) Partida con cuenta contable, subcuenta y auxiliares;
 - c) Comprobantes CFDI's;
 - d) Impuestos,
 - e) RFC,
 - f) Monto que ampara la póliza.

Metodología: Conceptos de pólizas

- Tipos de pólizas

- a) Póliza de diario
- b) Póliza de egreso
- c) Póliza de ingreso



Metodología: FactureApp creación de pólizas

- Definir la condición que permite obtener los CFDI que son de interés. Reglas del tipo:

SI (condición1) OP (condición2) OP (...)

- Asociar reglas a registros contables
- Aplicar las reglas para identificar los CFDI que las cumplen
- Genera las pólizas automáticamente

Metodología: Diseño RNN - datos

Variables de entrada:

- Método de pago
- Tipo de comprobante
- Forma de pago
- Uso del CFDI
- Clave de producto o servicio

Método de pago	Descripción
PUE	Pago en una sola exhibición
PPD	Pago en parcialidades o diferido

Tipo de comprobante	Descripción
I	Ingreso
E	Egreso
T	Nomina
P	Pago

Uso de CFDI	Descripción
G01	Adquisición de mercancías
G02	Devoluciones, descuentos o bonificaciones
G03	Gastos en general
I01	Construcciones
...	..

Forma de pago	Descripción
01	Efectivo
02	Cheque nominativo
03	Transferencia electrónica de fondos
04	Tarjeta de crédito
05	Monedero electrónico
06	Dinero electrónico
08	Vales de despensa
12	Dación de pago
13	Pago por subrogación
14	Pago por consignación
15	Condonación
17	Compensación
23	Novación
24	Confusión
25	Remisión de deuda
26	Prescripción o caducidad
27	A satisfacción del acreedor
28	Tarjeta de débito
29	Tarjeta de servicios
30	Aplicación de anticipos
31	Intermediario de pago
99	Por definir

Tabla 1 Valores para la variable forma de pago.

En el catálogo proporcionado por el SAT cuenta con 52 mil opciones

	División	Grupo	Clase	Subclase
Código	Primeros 2 dígitos	Siguientes 2 dígitos	Siguientes 2 dígitos	Últimos 2 dígitos
Ejemplo	50	20	17	06
Nombres	Alimentos, bebidas y tabacos	Bebidas	Café y té	Café molido

Metodología: Diseño RNN - datos

Variables de entrada:

Conjunto inicial de datos de dimensión 5 con 2226 registros

No	Método de pago	Forma de pago	Tipo de comprobante	Uso CFDI	Clave de producto o servicio
0	PUE	1	INGRESO	I02	01010101
1	PPD	99	INGRESO	S01	10101500
2	PUE	2	INGRESO	I05	10101500
2225	PUE	99	NOMINA	I04	10224700

Tabla 2 *Formato del registro de datos. La clave de producto o servicio se separó según la forma en la que se codifica. Por lo que la el registro tiene dimensión 8.*

Metodología: Diseño RNN - datos

Variables de salida (etiquetado):

El conjunto de datos consigna contempla 200 cuentas contables representadas por números consecutivos. No. se refiere al registro de entrada y (salida) a la cuenta contable a afectar

No.	y(salida
0	0
1	1
2	2
2225	37

Tabla 3 *Tabla de valores de salida deseados, de las cuentas contables a asignar.*

Metodología: Diseño RNN - datos

Representación de datos:

Variables con alta cardinalidad: Ahora la codificación de los datos se utilizó la técnica One-Hot (Hancock, 2020), donde se presentan todos los valores posibles para cada variable de entrada y se asigna valor 1 solo al valor correspondiente al de la entrada

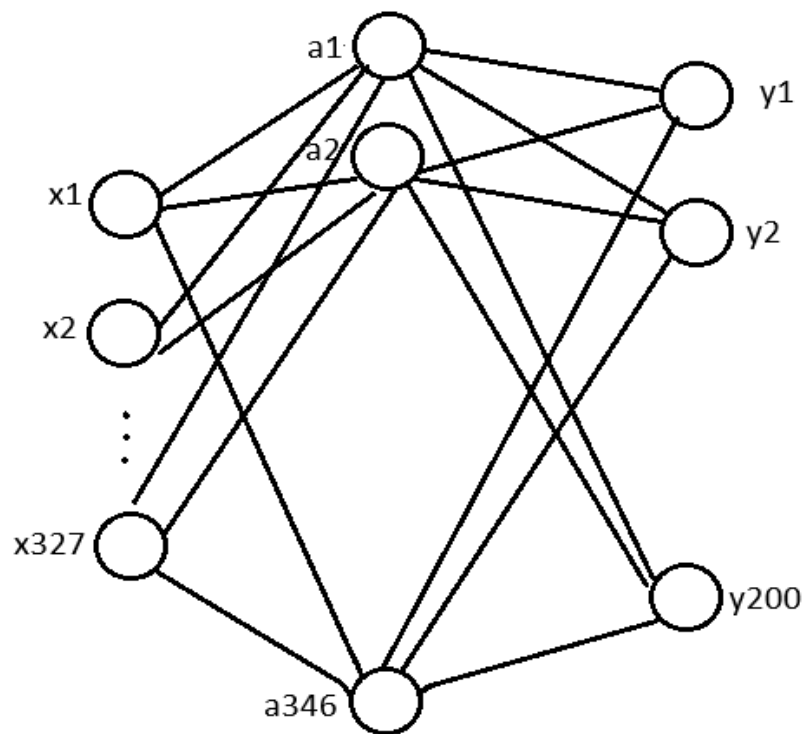
No.	0	1	2	3	...	324	325	326
0	1	0	1	0	...	0	0	0
1	0	1	0	0	...	0	0	0
...					...			
72130	1	0	0	0	...	0	0	0
72131	1	0	0	0	...	0	0	0

Dado que los datos para entrenar son pocos, se incrementó este conjunto de datos. Usando dos estrategias: duplicando aleatoriamente un registro y generando nuevos de manera artificial con datos conocidos hasta 71232.

Metodología: Diseño RNN - modelo

Modelo de red neuronal

Red neuronal con 327 entradas, una capa oculta de 346 neuronas y 200 de salida.



Keras de TensorFlow:

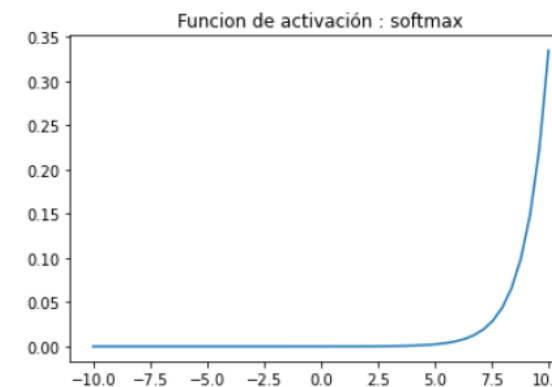
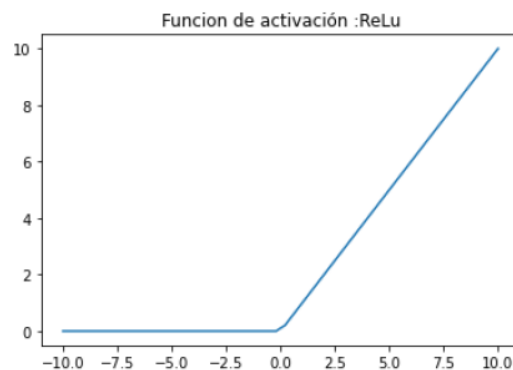
Capa oculta:

Función de activación ReLu

Regulador L2

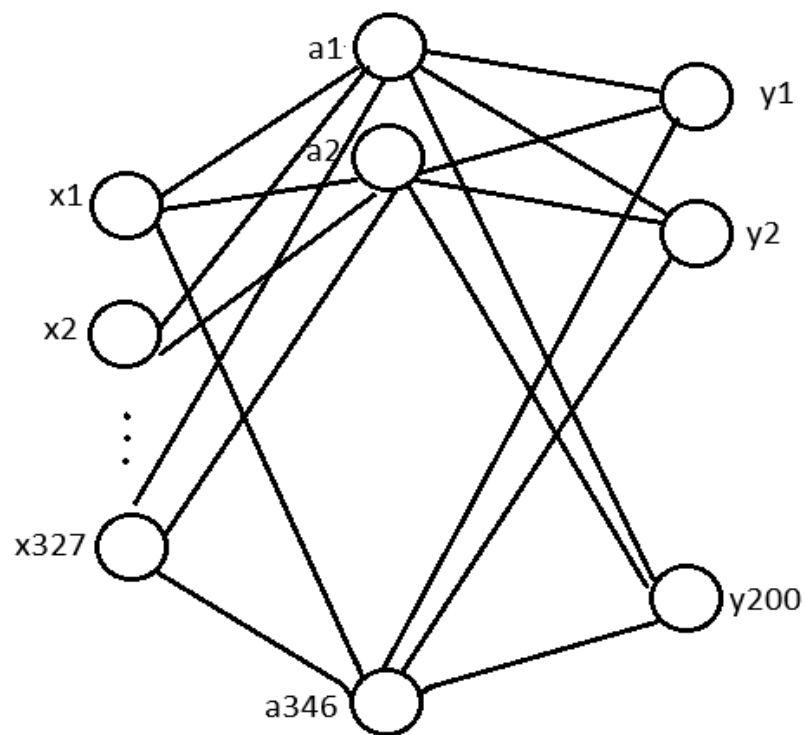
Capa de salida:

Función de activación SoftMax



Metodología: Diseño RNN - Entrenamiento

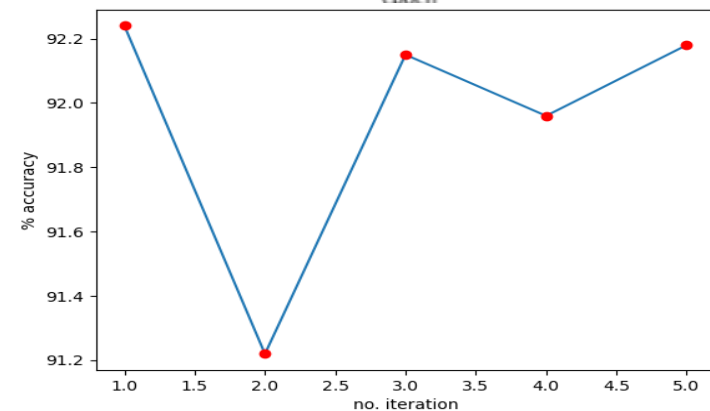
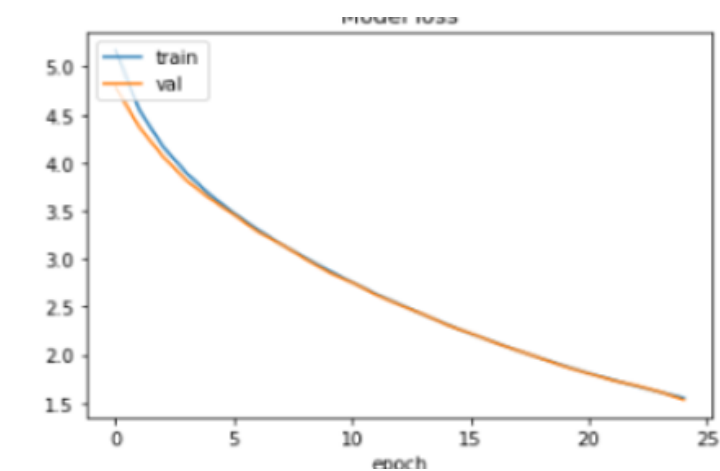
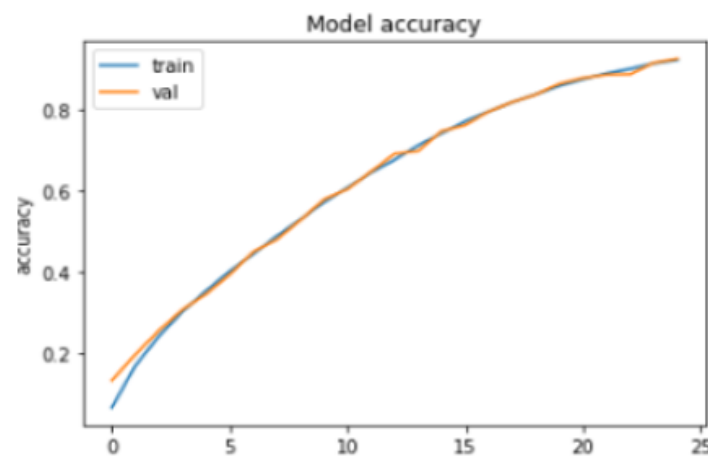
Modelo de red neuronal



Validación cruzada en 5 iteraciones

25 épocas por iteración

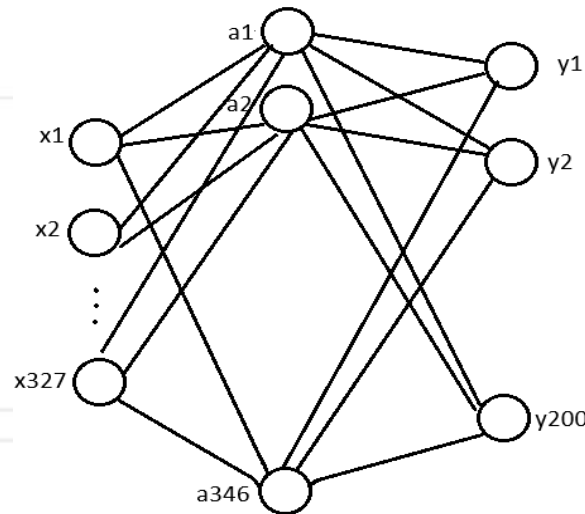
Datos 75% para entrenamiento y 25 para validación



Resultados

- Porcentaje de clasificación 92.18% de la red neuronal.
- REST API web para donde dado un CDFI sugiere las cuentas a afectar.

```
1  ... "data": {  
2  ... "metodoDePago": "PUE",  
3  ... "formaDePago": "1",  
4  ... "tipoDeComprobante": "INGRESO",  
5  ... "usoCFDI": "I02",  
6  ... "claveProdServicio": "01010101"  
7  ... },  
8  ... "numberRecommendations": 5  
9  
10
```



```
1  "prediction": [  
2  1,  
3  191,  
4  135,  
5  143,  
6  87  
7  ],  
8  "text": "Esta es la predicción..."  
9  
10
```

Conclusiones

- REST API para recomendaciones de cuentas que un CFDI puede modificar con un porcentaje de clasificación 92.18% basado en RNN.
- Análisis y codificación de los datos de un CFDI de interés
- Más experimentación con datos reales y usuarios finales es necesaria.
- Desarrollo de interfaces que permitan a los usuarios realizar consultas y recopilar datos que permitan entrenar la red o modificar el modelo.

Referencias

SAT. (2023) [https://www.sat.gob.mx/aplicacion/75169/servicio-de-facturacion-cfdi-version-4.0-\(vigente-a-partir-del-1-de-enero-de-2022\)](https://www.sat.gob.mx/aplicacion/75169/servicio-de-facturacion-cfdi-version-4.0-(vigente-a-partir-del-1-de-enero-de-2022))

Facture. (2023) <https://facture.com.mx/>

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. (2015) TensorFlow: Large-scale machine learning on heterogeneous.

Hancock, J.T., Khoshgoftaar, T.M. (2020) Survey on categorical data for neural networks. J Big Data 7, 28. <https://doi.org/10.1186/s40537-020-00305-w>



© RINOE-Mexico

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BECORFAN is part of the media of RINOE-Mexico., E: 94-443.F: 008- (www.rinoe.org/booklets)