

Clasificación automática de balances de empresas para la actualización de funciones de producción sectoriales de Chile

Joaquín Pérez Lapillo

Gerencia Estadísticas Macroeconómicas

Agenda

I. Motivación

- Importancia de las funciones de producción sectoriales
- Nuevas fuentes de información
- Nuevas herramientas analíticas

II. Metodología

- Compilación del set de entrenamiento
- Tratamiento del vector de texto
- Entrenamiento y selección de modelos
- Performance final

III. Resultados

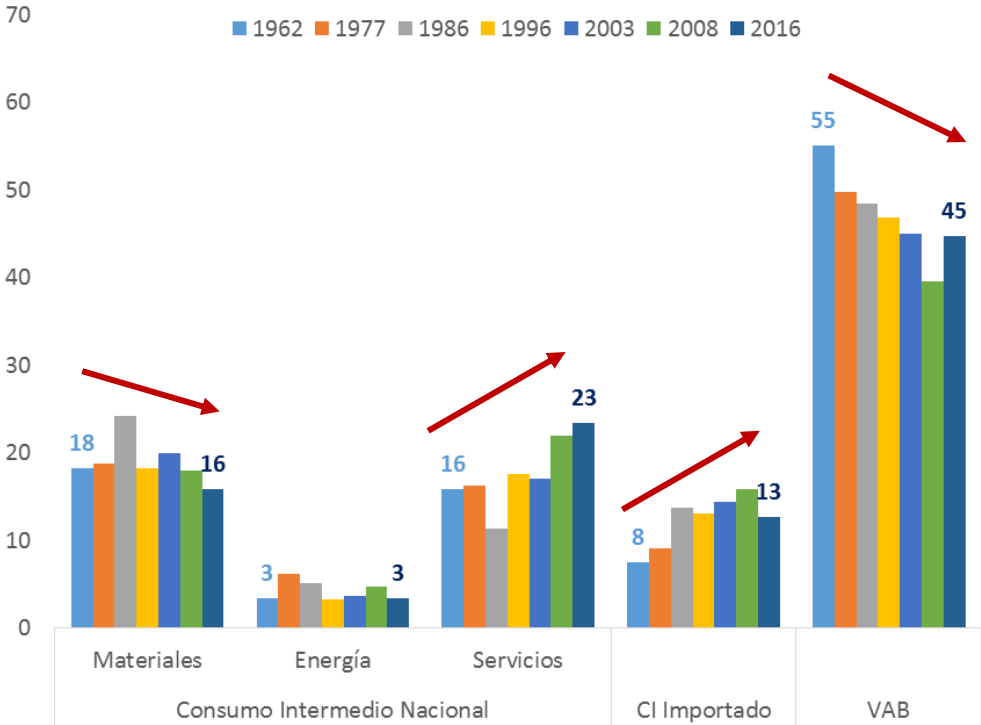
- Funciones de producción alternativas
- Discusión y conclusiones

A glass globe with a grid pattern, resting on a circular base. The globe is centered in the image. The background is a light blue grid with a faint, larger grid pattern overlaid. The text "I. Motivación" is positioned to the right of the globe, followed by a vertical orange bar.

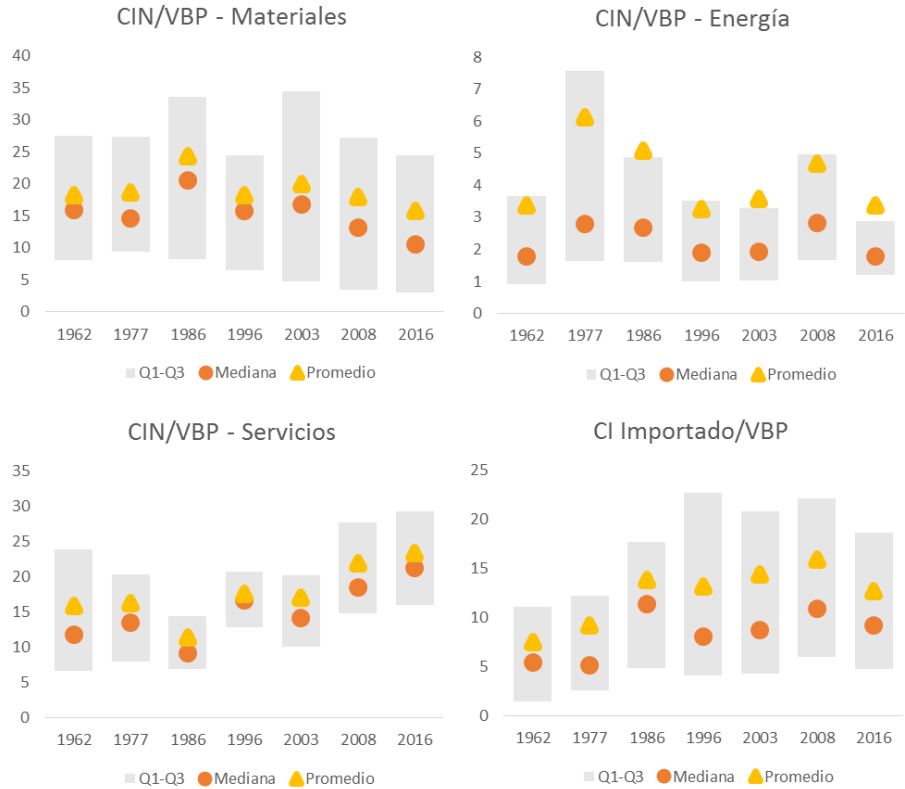
I. Motivación |

Relaciones insumo-producto actualizadas son claves para entender el funcionamiento de la economía.

Relaciones I-O, CIM y VAB
(% de la producción)



Dispersión sectorial por tipo de insumo
(% de la producción)



Fuente: BCCh en base a Matrices de Insumo-Producto (MIP).

Proyecto de actualización de funciones de producción se enmarca en la Compilación de Referencia 2018 de las Cuentas Nacionales de Chile.

Beneficios de contar con funciones de producción actualizadas

- Posibilita el **mejoramiento de las mediciones del PIB por actividad económica** y de las variables que componen el gasto, ahorro e inversión para el año de referencia.
- Proveen de un información vital para el **análisis estructural-descriptivo** de la economía (integración horizontal-vertical, cambios estructurales) y con ello la **calibración de modelos macro** de proyección.
- Por lo anterior, el presente trabajo se concentró en la determinación de la **estructura de consumo intermedio** a precios corrientes de cada actividad económica utilizando una **nueva fuente de información**.

Trabajo consistió en la explotación de una nueva fuente de información utilizando técnicas de aprendizaje automático.

Declaración Jurada (DJ) 1847 de Balances de empresas (SII)

- Información contable normalizada recibida por el SII, incluyendo glosas originales según el **plan de cuentas** de cada empresa.
- Considera empresas grandes y medianas, abarcando **80% de las ventas nacionales**.
- Para cada año tributario se reciben declaraciones de cerca de **25.000 empresas**.
- Trabajo se realizó sobre las bases correspondientes a los **años 2017 y 2018**:
 - En total, **367.000 partidas de costos (filas) por clasificar**
 - 7 ítems por empresa, por año, en promedio

Nuevas metodologías y software hacen posible su explotación

- Oportunidad de clasificar de forma automática las glosas de compras y gastos en bienes y servicios utilizando técnicas de **procesamiento de texto** y algoritmos de **aprendizaje automático** de tipo supervisados.
- Utilización de librerías disponibles en Python para **procesamiento eficiente**.

Trabajo consistió en la explotación de una nueva fuente de información utilizando técnicas de aprendizaje automático.

Id_Cuenta	Nombre_Cuenta	Debitos	Creditos	Saldo_Deudor	Saldo_Acreedor	Activo	Pasivo	Perdidas	Ganancias
3.01.03.00	GASTOS ADM OBRAS 700	9833409032	9452973...	380435794	0	0	0	380435794	0
3.01.02.00	COSTODEVENTAMDS.AFECTAS	358782159	0	358782159	0	0	0	358782159	0
3.01.02.00	COMISIONES	70909711	111501	70798210	0	0	0	70798210	0
3.01.02.00	COSTO DE VENTA	3460749488	0	3460749488	0	0	0	3460749488	0
3.01.02.00	MERCADERIAS E INSUMOS	51379970	5	51379965	0	0	0	51379965	0
3.01.03.00	ARRIENDOS	4200000	0	4200000	0	0	0	4200000	0
3.01.03.00	COMUNICACIONES	2078091	0	2078091	0	0	0	2078091	0
3.01.03.00	SEGUROS	381130	0	381130	0	0	0	381130	0
3.01.03.00	GASTOS VEHICULOS	266717	0	266717	0	0	0	266717	0
3.01.03.00	CONTRATO LEASING	3110931	3110931	0	0	0	0	0	0
3.01.03.00	MOVILIZACION Y TRASLADO	2103237	0	2103237	0	0	0	2103237	0
3.01.03.00	GASTOS VARIOS	2579969	0	2579969	0	0	0	2579969	0
3.01.03.00	FLETES Y CORREOS	143191	0	143191	0	0	0	143191	0
3.01.03.00	ELECTRICIDAD	26665	0	26665	0	0	0	26665	0
3.01.03.00	GAS Y CALEFACCION	39622	0	39622	0	0	0	39622	0
3.01.02.00	COSTO DE VENTA	35917991	0	35917991	0	0	0	35917991	0
3.01.02.00	TRANSPORTE	30083500	0	30083500	0	0	0	30083500	0
3.01.02.00	ELEMENTOS SEGURIDAD	7354505	0	7354505	0	0	0	7354505	0
3.01.02.00	OTROS GASTOS MATERIAS PRIMAS	707338	0	707338	0	0	0	707338	0
3.01.02.00	REPUESTOS MANTENCION	7370316	1140287	6230029	0	0	0	6230029	0
3.01.02.00	INSUMOS MATERIALES Y DETERGENTES	52390836	0	52390836	0	0	0	52390836	0
3.01.02.00	COMBUSTIBLE	9683867	0	9683867	0	0	0	9683867	0
3.01.02.00	COSTO VENTAS PRODUCTOS DEL GIRO	492695728	43665588	449030140	0	0	0	449030140	0
3.01.02.00	COSTO DIRECTO POR SERVICIOS	1833700	0	1833700	0	0	0	1833700	0
3.01.02.00	PRODUCTOS QUIMICOS Y DE FARMACIA	6700497	489124	6211373	0	0	0	6211373	0
3.01.02.00	PETROLEO	1051755	397765	653990	0	0	0	653990	0
3.01.02.00	BENCINA	2488669	1813976	674693	0	0	0	674693	0



II. Metodología

Proceso para la clasificación automática de glosas de costos de balances de empresas.



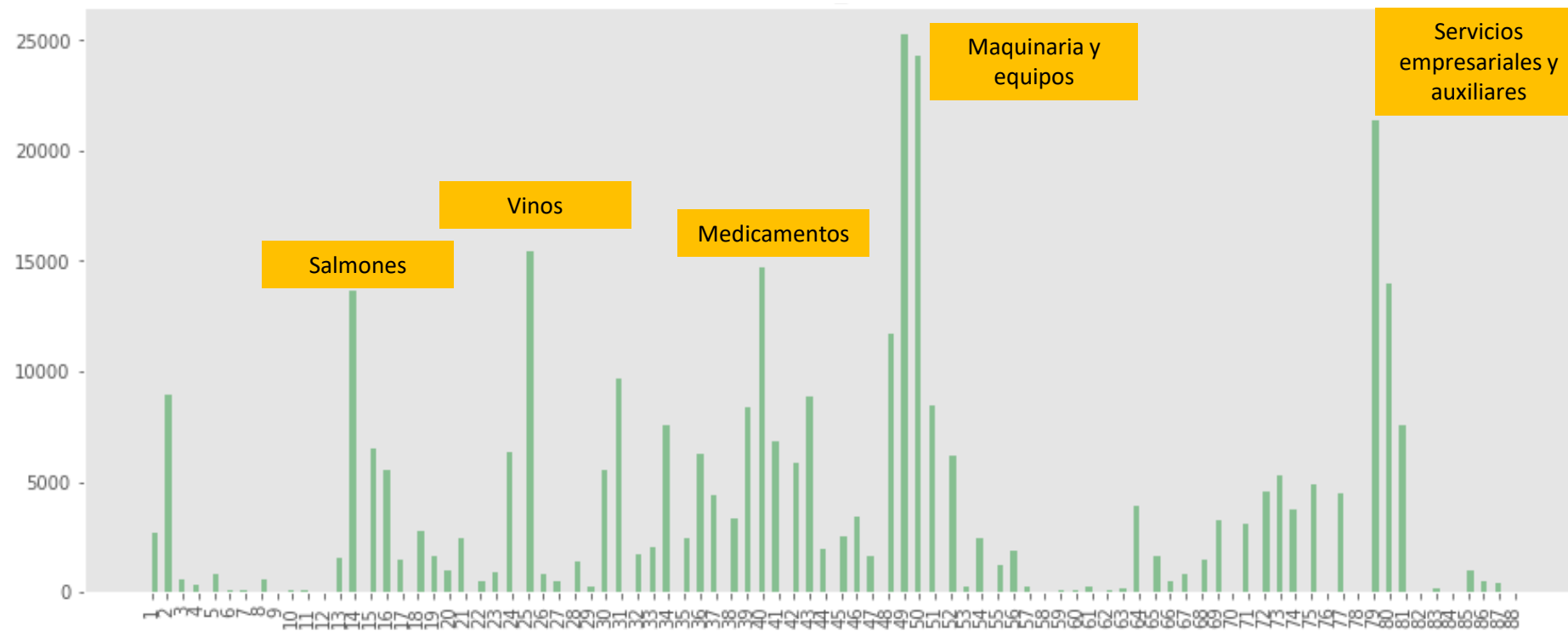
1. Consolidación de la base de entrenamiento.

- Estrategia consistió en consolidar 10 fuentes de datos para formar una base de entrenamiento que intenta **aproximar el universo de glosas de balances.**
- Cada fuente contiene glosas (contables y transaccionales) clasificadas de acuerdo al **Clasificador Único de Productos (CUP)** de Banco Central de Chile, en su extensión de **90 categorías.**
 - Similitud con reporte contable (combustibles, fletes, maderas)
 - Paridad con **nivel 2 CPC 2.1** (71 divisiones)
- Dado que cada base se especializa en ciertos productos, la base resultante presenta un importante **desbalance por clase.**

Fuente	N obs.
Declaraciones de exportación e importación (Aduanas)	163.462
Balances clasificados de empresas prioritarias	129.103
Encuesta Nacional de la Industria Anual (INE)	30.750
Web scrapping grandes supermercados	21.422
Clasificador único de productos BCCh	16.856
Compras públicas del gobierno (Chilecompra)	8.296
Encuestas BCCh de empresas de salud y educación	7.983
Canasta márgenes de comercio (INE)	4.000
CPC2 en inglés (UN)	2.997
Encuestas BCCh de empresas mineras	1.615
Total	386.484

Base de entrenamiento presenta desbalance significativo entre clases de productos.

Distribución base de entrenamiento por categoría CUP
(cantidad de elementos)



2. Estandarización del texto.

- Utilizando la librería NLTK se ejecuta:
 1. Remoción de **caracteres especiales** y dígitos.
 2. **Tokenización**: cada palabra en una frase se considera un elemento por si solo.
 3. Transformación a **minúsculas**.
 4. Remoción de **stopwords**: conectores se remueven de cada glosa (listado en español).
 5. **Stemming**: palabras son llevadas a su forma raíz.
 6. Filtro de tokens de **extensión menor a 3** caracteres.

ID_CUENTA	NOMBRE_CUENTA
3.01.02.00	SERVICIO DE CORRESPONDENCIA
3.01.02.00	HONORARIOS
3.01.02.00	COMUNICACIONES
3.01.02.00	AGUA
3.01.02.00	INSUMOS DE OFICINA
3.01.02.00	MANTENCION VEHICULOS
3.01.02.00	ARRIENDO
3.01.02.00	GASTOS COMUNES
3.01.02.00	CAPACITACION
3.01.02.00	SEGUROS
3.01.02.00	INSUMOS
3.01.02.00	COMBUSTIBLE
3.01.02.00	REPUESTOS
3.01.02.00	MATERIALES
3.01.02.00	REVISION TECNICA
3.01.02.00	MANTENCION MAQUINARIA
3.01.02.00	MANTENCION VEHICULOS
3.01.02.00	FLETES



Glosa
['servici', 'correspondent']
['honorari']
['comun']
['agu']
['insum', 'oficin']
['mantencion', 'vehicul']
['arriend']
['gast', 'comun']
['capacitacion']
['segur']
['insum']
['combust']
['repuest']
['material']
['revision', 'tecnic']
['mantencion', 'maquinari']
['mantencion', 'vehicul']
['flet']

3. Vectorización: *Word embeddings* y *TF*IDF*.

Word embeddings

- Es un grupo de técnicas desarrolladas por Google para **reducir la dimensionalidad** de un diccionario de tokens, accesible a través de la librería **Gensim** de Python.
- Con este método, cada token es transformado a un **vector numérico** continuo de “n” dimensiones.
- Comúnmente se realiza entrenando una red neuronal que aprende a representar de la manera deseada cada token, **reduciendo la dimensionalidad y manteniendo la distancia** entre el token y su vecindario.
- Además, existen **modelos previamente entrenados** disponibles para cada idioma.
- En este trabajo se utilizó un modelo ***FastText*** entrenado con 1.4 billones de palabras en español (y algunas en inglés) que transforma cada token en un **vector de 300 dimensiones**, disponible en:

<https://github.com/dccuchile/spanish-word-embeddings>

3. Vectorización: *Word embeddings* y *TF*IDF*.

```
# Print an example of the vector representation of a word
print(wv_model['arriend'])
```

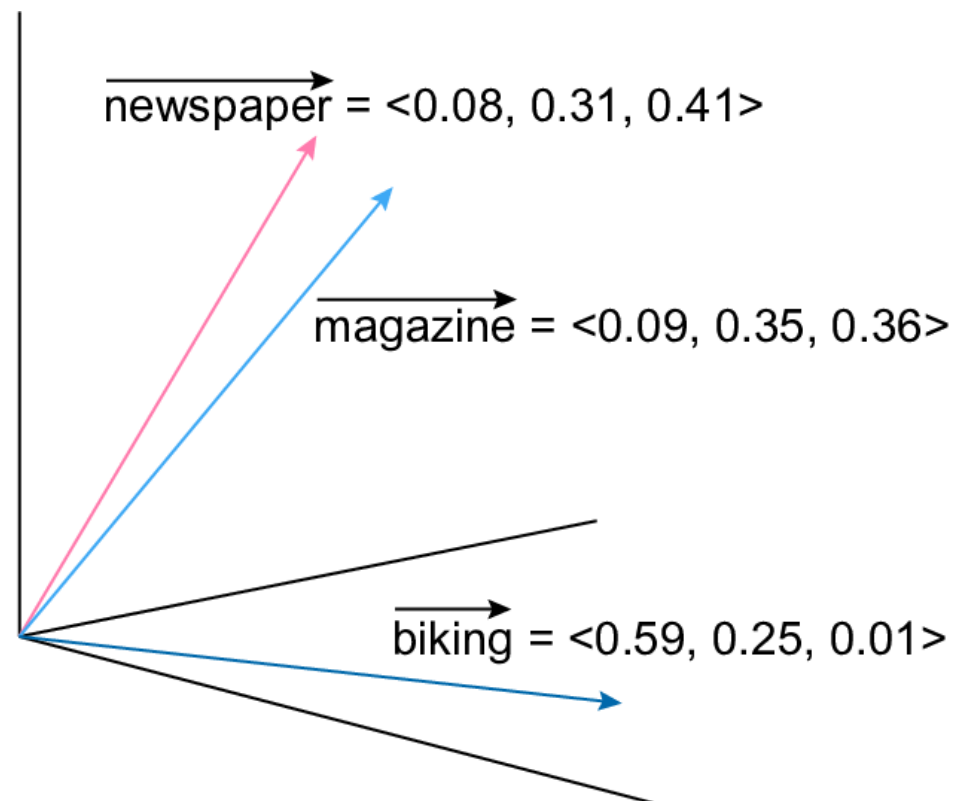
```
[-0.59  0.19 -0.55  0.03 -2.08  0.62 -0.78 -0.15  0.86 -0.61 -0.22 -0.34
 -0.81 -0.32  0.1  -0.31  0.13  0.35 -1.26 -0.09 -0.68 -0.15  1.4  0.3
  0.36  0.11  0.44  0.09  0.36 -0.49 -0.42  0.64 -0.7  1.17 -0.34  0.05
  0.71  0.15 -0.28 -0.71  0.54  0.6  0.31 -0.69 -0.81  1.01  0.15  0.07
  0.12  0.28 -0.08 -0.23 -0.29 -0.64 -0.59  1.16  0.71  0.24 -0.18 -1.19
  0.76  0.21 -0.22  1.41 -0.67 -0.33  0.92  1.2  0.31  0.35  0.52  1.21
 -0.16  0.25  0.2  -0.31 -0.36  0.21  0.86  0.33 -0.18 -0.29  0.36  1.16
  0.6  0.04 -0.36 -0.96 -1.03 -0.21 -1.18 -0.04  0.57 -0.97 -0.13 -0.24
  0.01  2.19 -0.4  -0.58]
```

```
# See the 10 most similar words
wv_model.wv.most_similar("arriend")
```

```
[('arrdo', 0.8899075984954834),
 ('arr', 0.871414065361023),
 ('alquil', 0.8285114765167236),
 ('manten', 0.7976149320602417),
 ('mantenc', 0.7887376546859741),
 ('repar', 0.7848337888717651),
 ('reparacion', 0.7780849933624268),
 ('arrend', 0.7770074009895325),
 ('subarr', 0.7648402452468872),
 ('mantencion', 0.7624949216842651)]
```

```
# We can analyse the model's performance trying different pairs of words
wv_model.wv.similarity(w1="arriend", w2="alquil")
```

0.8285115



Fuente: https://www.researchgate.net/figure/Visualization-of-3-dimensional-word-embeddings_fig1_325451970

3. Vectorización: *Word embeddings* y *TF*IDF*.

*TF*IDF: Term frequency / Inverse document frequency*

- Luego de que cada token logra su representación numérica, cada fila de nuestra base de entrenamiento es transformada a un **vector numérico** usando **TF*IDF**.
- TF*IDF **pondera palabras según su relevancia** (poco común) en el total de documentos disponible:

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

- Finalmente, combinado Word embeddings y TF*IDF es posible representar numéricamente cada fila:

Glosa
['servici', 'correspondent']
['honorari']
['comun']
['agu']
['insum', 'oficin']
['mantencion', 'vehicul']
['arriend']
['gast', 'comun']



```
array([[ 1.3057307 , -0.08263066,  0.83302128, ..., -0.85772377,
         5.04536533, -0.67781782],
       [ 0.8528434 , -2.39659214, -1.48367858, ..., -0.71295035,
         5.06377888, -2.91519928],
       [-1.62950683, -1.42616856,  2.44736338, ..., -4.61831999,
         2.08979034, -7.00978184],
       ...,
       ...])
```

4. Entrenamiento de algoritmos.

- Bajo el paradigma de **aprendizaje supervisado**, se entrenan 5 tipos de algoritmos clásicos de *machine learning*, usando la librería Sklearn de Python.
- Para testear la calidad de la predicción, aislamos un **20% de las observaciones** de la base de entrenamiento.
- Se seleccionan los mejores parámetros para los algoritmos seleccionados mediante **grid search** aplicando validación cruzada (k=5).
- Antes del entrenamiento, la presencia de cada clase en la **base de entrenamiento se balancea** mediante *up/down sampling* de manera tal de reducir posibles sesgos.
- Algoritmos seleccionados y sus mejores parámetros de acuerdo a **grid search**:
 - **Logistic regression (LR)**: multiclass
 - **Random forest (RF)**: 300 árboles, profundidad máxima 50, entropía
 - **Naive Bayes (NB)**: Gaussian kernel
 - **Multilayer perceptrón (MLP)**: 2 capas de 200 neuronas cada una, Adam optimizer
 - **Support vector machine (SVM)**: Polynomial kernel, order 3, C=10

5. Evaluación de performance *in sample*.

- Selección se realiza en base a performance en términos de **accuracy** (total de aciertos) y **F1-score** (promedio entre *precision* y *recall*, por clase) sobre el 20% de los datos que no se mostró en el training.
- **Random forest** supera al resto en términos de accuracy, mientras **SVM** presenta un mayor F1 score.
- Sin embargo, tanto en entrenamiento como predicción SVM tarda **3.5 veces más** que RF, lo que implica un **alto costo computacional** sin un claro beneficio en performance.
- Además, siendo RF un modelo probabilístico provee **medidas de robustez** de la predicción.
- De todas formas, elección de modelo final depende de performance sobre la **base objetivo a clasificar**.

Model	Accuracy	F1-Score
LR	57.5	44.6
RF	76.2	69.5
NB	41.6	32.5
MLP	66.0	56.5
SVM	73.0	74.9

6. Evaluación final (*out of sample*).

- Glosas de balances se estandarizan siguiendo los pasos descritos para la base de entrenamiento (puntuación, tokenizer, stemming, etc.).
- Luego, se utiliza el modelo de **Word embeddings** entrenado previamente y **TF*IDF** para obtener la representación numérica de la base, considerando las mismas dimensiones (300).
- Los atributos resultantes son **entregados a los modelos** de RF y SVM, que entregan una predicción (CUP) para cada glosa.
- Mediante 10 muestras aleatorias de 100 observaciones cada una, **expertos evaluaron manualmente** la performance de ambos modelos.
- Niveles de **precisión final favorecen a RF** por poco margen.

Model	Accuracy
Random forest	67.2
Support vector machine	66.8

6. Evaluación final (*out of sample*).

PERIODO	RF	cae_n1_2013	ID_CUENTA	NOMBRE_CUENTA	PERDIDAS	CONCEPTO_PARTIDA	Filter
0	201900	40016	69.0	3.01.02.00	COSTOS ADUANA	1620633	5 0
1	201800	16017	65.0	3.01.02.00	SEGUROS	2045501	5 0
2	201800	16017	65.0	3.01.02.00	COMBUSTIBLE	10607380	5 0
3	201800	16017	65.0	3.01.02.00	REPUESTOS	11945070	5 0



Glosa	
0	[cost, aduan]
1	[segur]
2	[combust]
3	[repuest]



```
array([[ 1.3057307 , -0.08263066,  0.83302128, ..., -0.85772377,
         5.04536533, -0.67781782],
        [ 0.8528434 , -2.39659214, -1.48367858, ..., -0.71295035,
         5.06377888, -2.91519928],
        [-1.62950683, -1.42616856,  2.44736338, ..., -4.61831999,
         2.08979034, -7.00978184],
        ...,
        ...])
```



PERIODO	RF	cae_n1_2013	ID_CUENTA	NOMBRE_CUENTA	PERDIDAS	CONCEPTO_PARTIDA	Filter	CUP_89	PercVote
0	201900	40016	69.0	3.01.02.00	COSTOS ADUANA	1620633	5 0	71	11.52
1	201800	16017	65.0	3.01.02.00	SEGUROS	2045501	5 0	75	90.96
2	201800	16017	65.0	3.01.02.00	COMBUSTIBLE	10607380	5 0	38	72.22
3	201800	16017	65.0	3.01.02.00	REPUESTOS	11945070	5 0	51	40.93

Código de producto (predicción)

RF permite obtener el % de votos con los cuales se asignó la predicción

6. Evaluación final (*out of sample*).

Aciertos (67%); Prom. votos RF 48.6%

NOMBRE_CUENTA	CUP_89	N_CUP	PercVote
ENERGIA ELECTRICA	54	Distribución y transmisión eléctrica	100,0
FLETES Y TRANSPORTES	69	Transporte de carga	100,0
ARRIENDO EMBARCACIONES	70	Servicios de alquiler de medios de tr	100,0
ARRIENDOS PAGADOS	77	Servicios inmobiliarios	100,0
ARRIENDO DE VEHICULOS	80	Servicios administrativos y de apoyo	100,0
GASTOS BANCARIOS	74	Servicios financieros no SIFMI	99,4
ARRIENDO DE OTROS BIENES MUEBLES	80	Servicios administrativos y de apoyo	99,0
AGUA	56	Agua	97,4
MANTENCI N INSTALACIONES	64	Servicios especializados de construc	97,4
SERVICIOS DE CORRESPONDENCIA	71	Actividades conexas de transportes	96,7
TAG	71	Actividades conexas de transportes	96,0
OTRAS ASESORIAS	79	Servicios profesionales, científicos y	95,9
TELEFONOS	72	Comunicaciones	93,6
COMPRA DE GAS	55	Gas	90,3
SEGUROS - INCENDIO	75	Servicios de seguros	89,0
SAL	13	Otros minerales	87,3
ENERGIA ELECTRICA OPERACIONES	54	Distribución y transmisión eléctrica	84,4
COMISIONES	65	Servicios comerciales	75,3
LEASING	74	Servicios financieros no SIFMI	73,1
MANTENCIÓN EQUIPOS CONTRATOS	81	Servicios de reparaciones	65,3
FLETES Y ALMACEN.IMPORTACION	69	Transporte de carga	41,8
ARRIENDO OFICINAS Y BA OS MODULARE	77	Servicios inmobiliarios	40,9
MANTENCIÓN OFICINAS. EDIFICIO	64	Servicios especializados de construc	35,9
CORREOS Y VALIJAS ASEG. CALIDAD	71	Actividades conexas de transportes	24,7
SERVICIOS DE MANTENCIÓN	80	Servicios administrativos y de apoyo	24,3
TRANSPORTE ESCOLAR	68	Transporte de pasajeros	24,1
ENERGIA RIEGO	54	Distribución y transmisión eléctrica	23,6
SERVICIOS PROFESIONALES EXTERNALIZA	79	Servicios profesionales, científicos y	18,6

Errores (33%); Prom. Votos RF 12.3%

NOMBRE_CUENTA	CUP_89	N_CUP	PercVote
RASTREO SATELITAL	74	Servicios financieros no SIFMI	87,7
LABORATORIO	64	Servicios especializados de construc	63,8
ARRIENDO ALZAHOMBRE ARTICULADO	77	Servicios inmobiliarios	21,1
COSTO ABARROTES	69	Transporte de carga	13,1
CONTRATISTA PERSONAL EN PLANTA	64	Servicios especializados de construc	11,9
GASTOS POSTVENTA OBRAS	79	Servicios profesionales, científicos y	11,1
SERVICIOS DE TERCEROS ENCUADERNACI	80	Servicios administrativos y de apoyo	10,4
UTILES DE OFICINA	81	Servicios de reparaciones	9,0
COSTO SUB CONTRATOS	69	Transporte de carga	9,0
MANT.VEHICULO FORESTAL LOS TILOS LC	4	Productos silvícolas	9,0
VENTAS NACIONAL CLIENTE FINALMANU	79	Servicios profesionales, científicos y	8,7
ATENCION INVITADOS	80	Servicios administrativos y de apoyo	8,2
COSTO POR EVENTOS	79	Servicios profesionales, científicos y	8,0
COSTOS DE EXPLOTACI N	80	Servicios administrativos y de apoyo	7,9
CIELO PANEL	50	Maquinaria y equipo eléctrico y elec	7,3
GASTOS OBRA ARQUITECTURA LAS COND	77	Servicios inmobiliarios	7,0
PREPARACION DE SUELO	41	Otros productos químicos	6,8
MATERIALES GENERALES	81	Servicios de reparaciones	6,6
COSTO M.PRIMAS Y MATERIALES	75	Servicios de seguros	6,5
COS STND: SMALL PROJ LABOR	73	Servicios de información	6,4
COSTOS ESTIMADOS	65	Servicios comerciales	6,3
ARMADURA CADENA 2Âº PISO MATERIAL	30	Textiles	6,3
BENEFITS - INSURANCE - DENTAL - OPS	73	Servicios de información	6,3
TASAS SOBREVUELOMAN	80	Servicios administrativos y de apoyo	5,9
M.PRIMA Y MATERIALES CROM	45	Otros productos minerales no metál	5,6
PT IMPORT BOMBIN PREMIUM	73	Servicios de información	5,4
COSTO MAT.PRIMAS RANCHO LA HERRAC	65	Servicios comerciales	5,0
DAP CAMPANA M EXT PN D365	40	Productos farmacéuticos y de higien	5,0

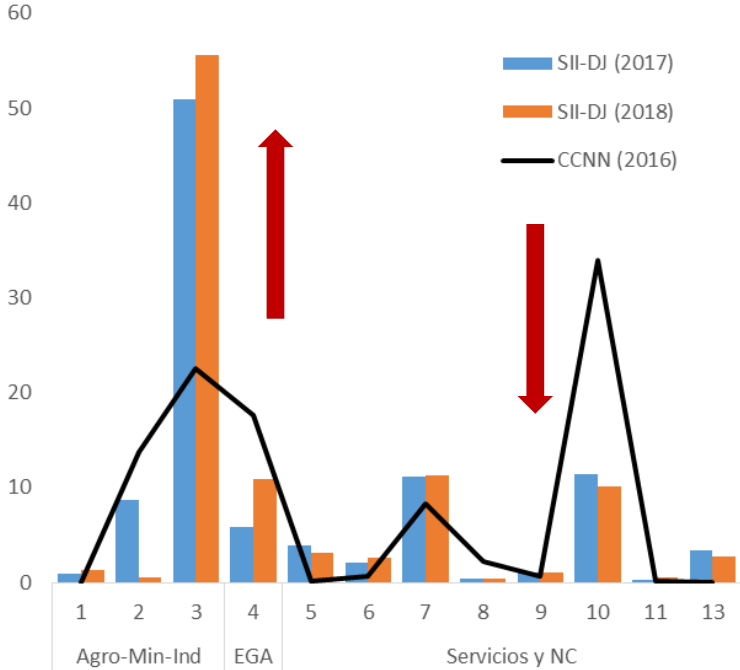
A line chart with a light blue grid. The y-axis is labeled 'visits' and has a value of 1000. The x-axis has values 14, 16, 18, 20, 22, 24, 26, 28. There are two data series: a blue line and a brown line. The blue line starts at approximately 1000 at x=14 and fluctuates upwards to about 1500 at x=28. The brown line starts at approximately 500 at x=14 and fluctuates upwards to about 1000 at x=28. A light brown shaded area is present between the two lines from x=14 to x=20. The text 'III. Resultados' is overlaid on the chart in a dark blue font, followed by a vertical brown bar.

III. Resultados

Al comparar con estadísticas oficiales, resultados proponen distinta intensidad de uso de algunos insumos.

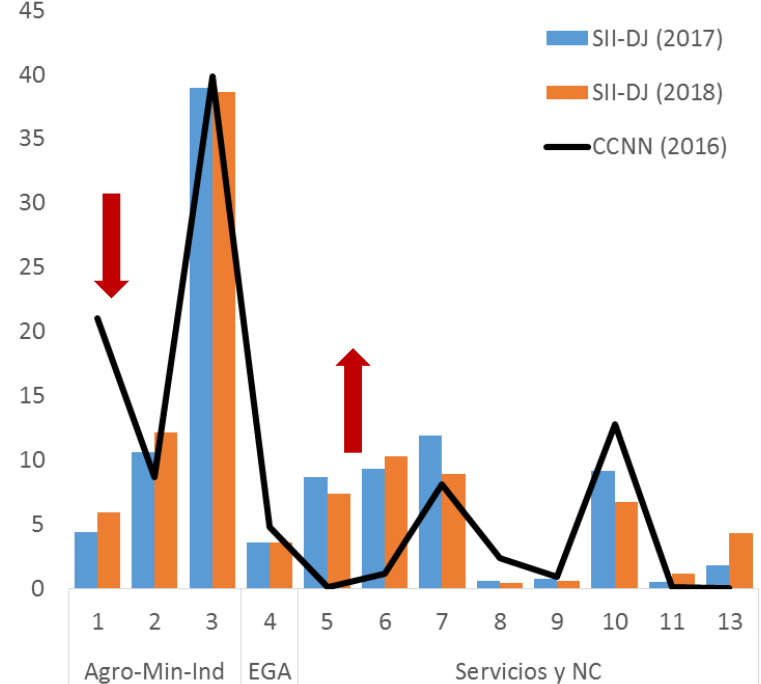
Minería

(% del consumo intermedio)



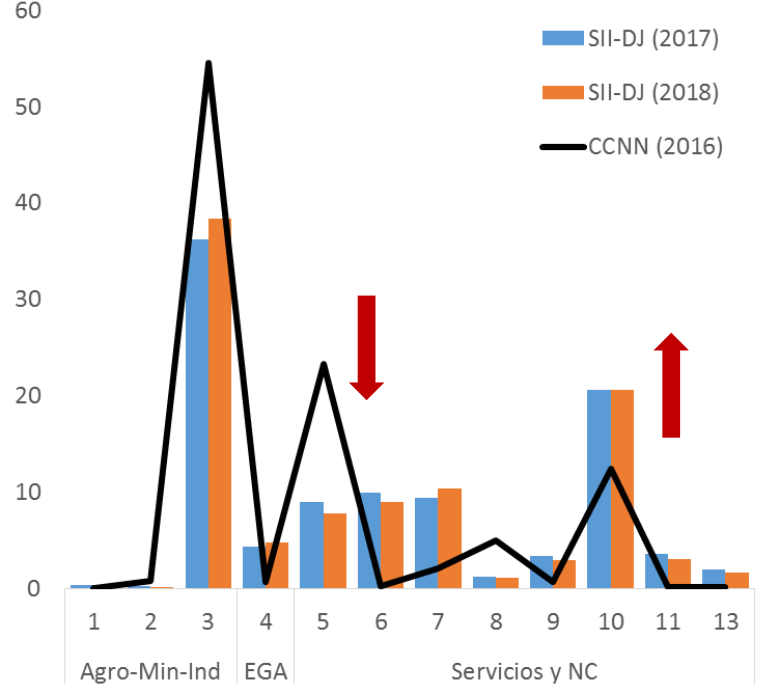
Manufactura

(% del consumo intermedio)



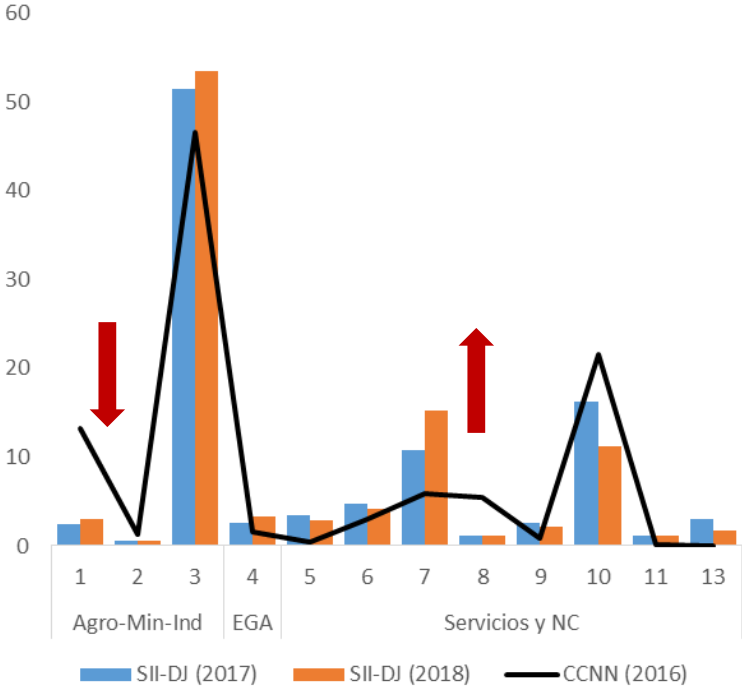
Construcción

(% del consumo intermedio)

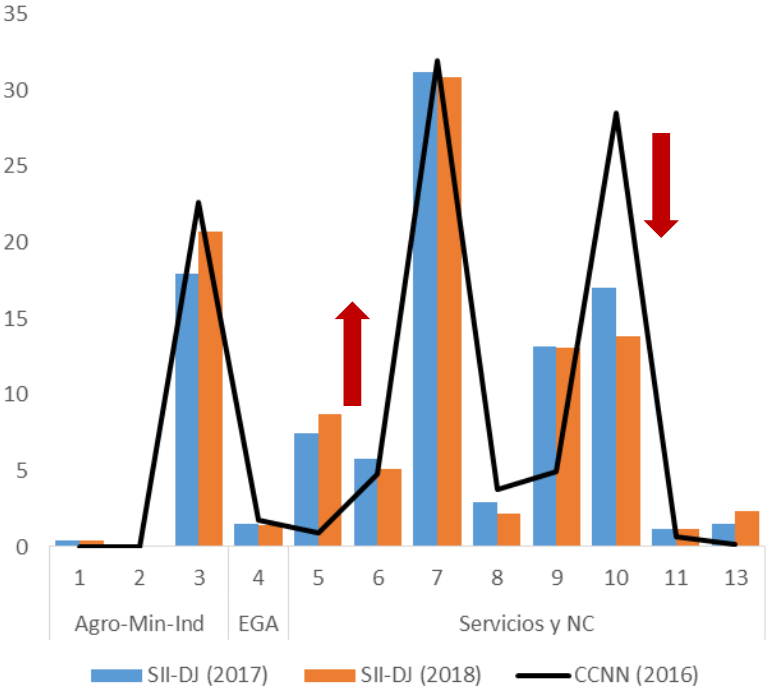


Al comparar con estadísticas oficiales, resultados proponen distinta intensidad de uso de algunos insumos.

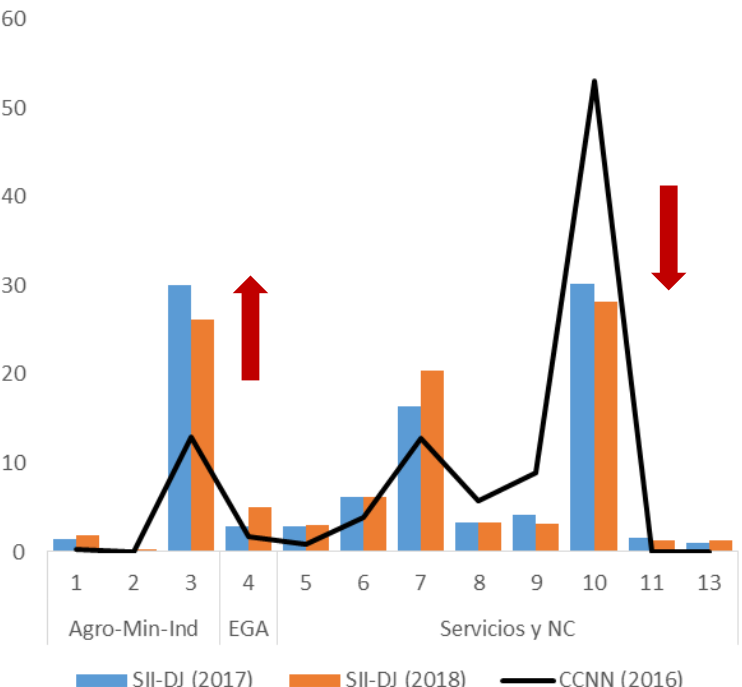
Agropecuario-silvícola y pesca (% del consumo intermedio)



Transporte, información y com. (% del consumo intermedio)

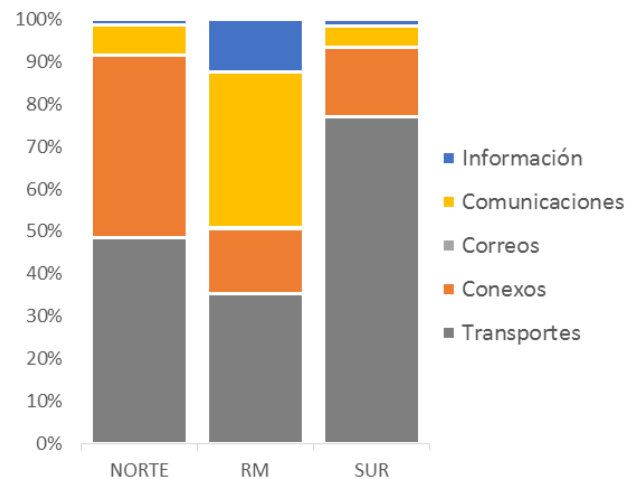
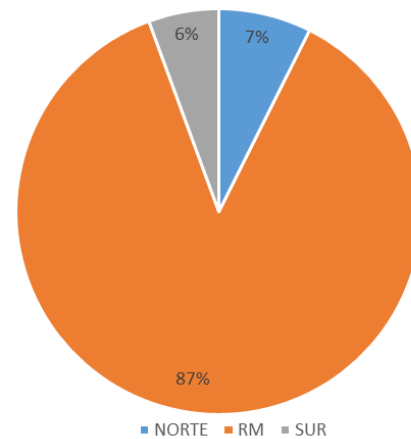
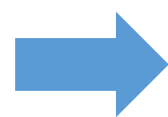
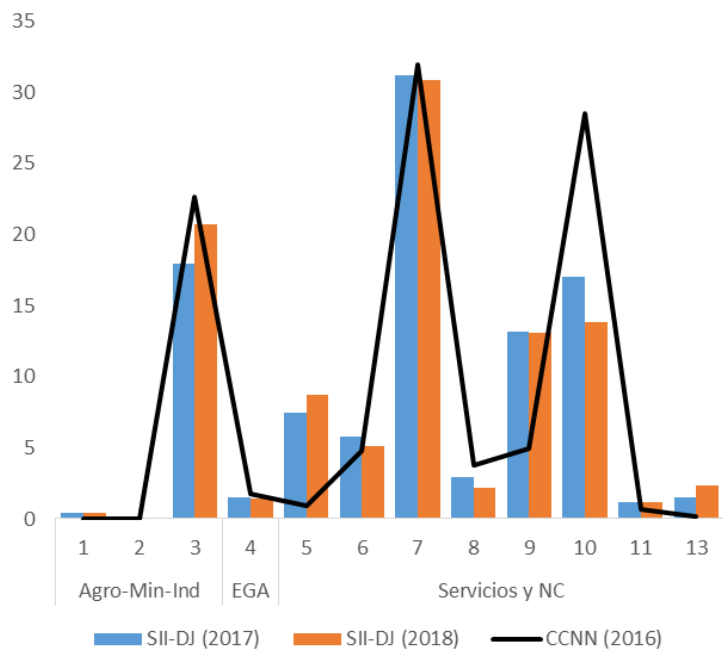


Servicios empresariales (% del consumo intermedio)

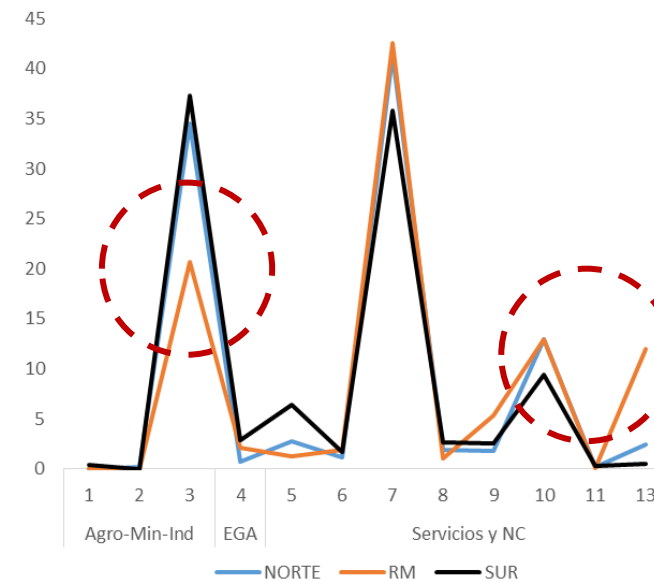


Base a nivel de microdatos permite analizar estructuras diferenciadas por zona geográfica.

Transporte, información y comunic.
(% del consumo intermedio)




Transporte
(% del consumo intermedio)



Discusión y próximos pasos.

- En general, performance obtenida por los mejores algoritmos entrenados alcanza **niveles aceptables** para un **problema real**.
- Además, proceso implicó un **bajo costo** ya que **no se invirtieron recursos** en construir una base de entrenamiento ad-hoc al problema (estrategia ideal).
- La base de datos de costos clasificados resultante sirve hoy como **contraste para la elaboración de funciones de producción sectoriales**, en el marco de la Compilación de Referencia 2018 de Chile.
- El modelo clasificador de glosas contables a productos ya ha sido utilizado con **otros fines internos**, tales como la elaboración del vector de inventarios.
- Actualmente, se encuentra en proceso la **adaptación del *pipeline*** a otro tipo de datos administrativos de **tipo transaccional** (boletas y facturas electrónicas diarias), lo que implica desafíos desde el punto de vista del volumen de información a clasificar.



Clasificación automática de balances de empresas para la actualización de funciones de producción sectoriales de Chile

Joaquín Pérez Lapillo

Gerencia Estadísticas Macroeconómicas