

---

# SISTEMA RECOMENDADOR PARA CURSOS DE FORMACIÓN COMPLEMENTARIA EN EL CENTRO DE TELEINFORMÁTICA Y PRODUCCIÓN INDUSTRIAL DEL SENA REGIONAL CAUCA.

Jorge Andrés Coral Trujillo  
Ingeniero de Sistemas  
Instructor Desarrollo de Software  
SENA - Regional Cauca  
jorgect77@misena.edu.co

Pablo Antonio Ortiz Gutiérrez  
Ingeniero de Sistemas  
Instructor Desarrollo de Software  
SENA - Regional Cauca  
ortizpablos@misena.edu.co

Asesor de Proyecto  
Cristhian Figueroa  
Doctor en Ingeniería Telemática  
Docente Universidad del Cauca  
cfigmart@unicauca.edu.co

## RESUMEN

En el presente documento se plantea un sistema de recomendación utilizando el filtrado colaborativo basado en el comportamiento de usuarios, al tomar los datos de los aprendices del Servicio Nacional de Aprendizaje "SENA", que cursaron un programa de formación complementaria virtual o presencial.

Aplica el coeficiente de correlación de Pearson para el cálculo de similitud entre ítems, al tomar los diferentes estados del aprendiz como: **Certificado = 4**, terminó el curso y el sistema generó su certificado. **Por certificar = 4**, terminó el curso con éxito y está por ser certificado. **En Formación = 3**, está en el proceso de formación. **Retiro Voluntario = 2**, el aprendiz decide retirarse del curso. **Cancelado = 1**, el aprendiz abandona el curso.

Para el desarrollo del sistema de recomendación se empleó la metodología CRISP-DM la cual organiza cada fase dentro del ciclo de vida del proyecto desde el entendimiento del negocio hasta su evaluación y despliegue.

Creado el modelo, identifica un usuario con diferentes estados dentro de la plataforma; se entrena y da como proceso final resultados de recomendación dentro de la misma línea de conocimiento.

**Palabras clave:** Inteligencia Artificial, Machine Learning, Sistema Recomendador, Filtrado Colaborativo

## 1. INTRODUCCIÓN

Con el avance de la tecnología el área de la educación ha tenido que adaptarse a esta inevitable tendencia y avanzar de acuerdo con las novedades y nuevas tendencias del sector como el uso de machine learning o deep learning, relacionadas con la Inteligencia Artificial, que impulsan la gestión del conocimiento [1].

En la actualidad existen gran variedad de sistemas de recomendación que ayudan al usuario o cliente en la toma de decisiones, por ejemplo: Netflix que recomienda series o películas de acuerdo con el contenido que haya visto el usuario, Spotify que recomienda música para el usuario de acuerdo con lo que haya escuchado anteriormente.[2]

Teniendo en cuenta que la misión del Servicio Nacional de Aprendizaje SENA, "está encargado de cumplir la función que le corresponde al Estado de invertir en el desarrollo social y técnico de los trabajadores colombianos, ofreciendo y ejecutando la formación profesional integral, para la incorporación y el desarrollo de las personas en actividades productivas que contribuyan al

desarrollo social, económico y tecnológico del país”.[3]

La mayor demanda en el SENA tiene que ver con la formación complementaria por su corta duración, permitiendo la actualización o desarrollo de habilidades y destrezas que corresponden a temáticas específicas del sector productivo.

La variedad y cantidad de cursos ofertados, complica la búsqueda de estos; por esta razón las empresas e instituciones están incorporando herramientas que permitan ofrecer o recomendar temáticas y productos de acuerdo con el interés o afinidades del usuario. Una de las más actuales y conocidas es el sistema recomendador, conjunto de técnicas y herramientas que generan sugerencias útiles para el usuario de acuerdo con sus gustos.

Con base en lo anterior se plantea un sistema de recomendación que le permita al usuario seleccionar una de las sugerencias brindadas; el sistema toma la información del usuario según cursos de formación complementaria realizados en el Servicio Nacional de Aprendizaje SENA.

Existen varias técnicas para los sistemas de recomendación y de acuerdo con los intereses del usuario se decide optar por la técnica de Filtrado Colaborativo, que tiene que ver con las recomendaciones que recibe un usuario teniendo en cuenta otros usuarios con similares preferencias. Los gustos similares entre usuarios se calculan según la similitud entre calificaciones o puntuación de los ítems.[4].

Los métodos basados en vecinos más cercanos tienen gran popularidad por su simpleza, eficiencia y habilidad para hacer recomendaciones más precisas y personalizadas.[4]

Para el desarrollo del proyecto se definen las siguientes fases:

**Materiales y métodos:** describe la fuente de la información a trabajar y el manejo que le dará a la misma, además se selecciona la metodología que lleva a cabo el ciclo de vida del proyecto.

**Resultados:** para comprobar y evaluar la exactitud de los datos obtenidos y determinar qué tan confiable es la recomendación brindada por el sistema.

**Discusión:** se realiza el análisis de las recomendaciones obtenidas para cada usuario con el que se prueba el sistema.

## 2. MATERIALES Y MÉTODOS

En el desarrollo del proyecto se implementa la metodología “CRISP-DM (Cross Industry Standard Process for Data Mining), es modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema”. [5].

La metodología CRISP-DM estructura el ciclo de vida de un proyecto en seis fases, que se relacionan entre ellas. Puede ser estudiada, modificada, y utilizada libremente. [5].

En esta propuesta se seguirá cada una de las fases dadas por la metodología CRISP-DM, la finalidad será determinar los objetivos, requisitos y evaluar la situación actual del proceso. El ciclo de vida de la metodología se muestra en la figura 1.

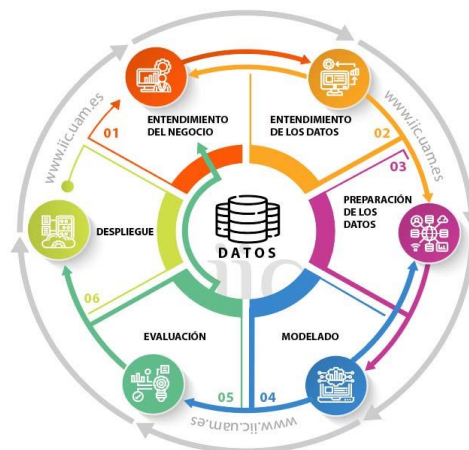


Figura 1. Esquema del ciclo CRISP-DM estándar.

### 2.1. Comprensión del Negocio.

El SENA, tiene oferta educativa de formación a nivel nacional en diferentes niveles como: operarios, cursos complementarios, técnicos, tecnólogos y especializaciones tecnológicas de modalidad presencial o virtual.

Los aspirantes una vez matriculados a un programa de formación en cualquier nivel su rol cambia a aprendiz quien tiene acceso a plataformas para el proceso de gestión académica y administrativa como [www.senasofiaplus.edu.co](http://www.senasofiaplus.edu.co), durante su formación el

aprendiz tiene diferentes estados: **Certificado:** termina y aprueba la formación y el sistema le genera el certificado. **Por certificar:** termina y aprueba con éxito su formación y está pendiente por generar el certificado. **Formación:** está en proceso de estudio. **Retiro Voluntario:** abandona la formación. **Cancelado:** deserta de la formación.

## 2.2. Comprensión de los datos.

El Servicio Nacional de Aprendizaje SENA, cuenta con la plataforma SOFIAPLUS (Sistema Optimizado para la Formación y el Aprendizaje Activo), que opera como la principal herramienta para la gestión de los procesos formativos; la plataforma, en el rol de instructor, permite la descarga de diferente información como el reporte **DF\_14-Matriculados Detallado**, con información del aprendiz como muestra el siguiente consolidado en la tabla 1.

REPORTE	CAMPO	DESCRIPCION	TIPO DE DATO
DF_14-Matriculados Detallado	CODIGO_REGIONAL	Código que identifica la regional Cauca	int64
	REGIONAL	Identifica el nombre de la regional	object
	CODIGO_SEDE	Identifica el código del centro de formación	int64
	SEDE	Identifica el nombre del centro de formación	object
	TIPO_DOCUMENTO	Identifica el tipo del documento del Aprendiz	object
	NUMERO_DOCUMENTO	Identifica al número de documento del Aprendiz	object
	NOMBRE	Describe el nombre o nombres del aprendiz	object
	PRIMER_APELLIDO	Describe el primer apellido del aprendiz	object
	SEGUNDO_APELLIDO	Describe el segundo apellido del aprendiz	object
	ESTADO_APRENDIZ	Describe el estado en que se encuentra del aprendiz	object
	CONVENIO	Identifica el nombre del convenio con otra institución	object
	TIPO_DOCUMENTO_EMPRESA	Identifica el tipo de documento de la empresa	float64
	NUMERO_DOCUMENTO_EMPRESA	Identifica el número de la empresa	float64
	EMPRESA	Describe el nombre de la empresa	float64
	FICHA	Identifica el número del curso	int64
	ESTADO_FICHA	Describe el estado de la ficha	object
	CODIGO_PROGRAMA	Identifica el número del programa de formación	int64
VERSION_PROGRAMA	Identifica la versión de programa de formación	int64	
NIVEL_DE_FORMACION	Describe el nivel del curso	object	
PROGRAMA	Describe el nombre del programa de formación	object	
IDENTIFICADOR_CONVENIO	Identifica el número de convenio con otra institución	int64	
AMPLIACION_COBERTURA	Describe si pertenece o no a la estrategia de cobertura	object	

Tabla 1. Consolidado DF\_14-Matriculados Detallado.

Se realiza el análisis exploratorio de los datos con el fin de conocer la estructura general y las propiedades de estos que puedan ser de utilidad creando tablas de frecuencia y gráficos para generar un informe de los datos obtenidos.

Los dos cursos de mayor demanda entre los usuarios son Sistemas y English dot works 1 (inglés) como muestra la siguiente tabla 2. de frecuencia.

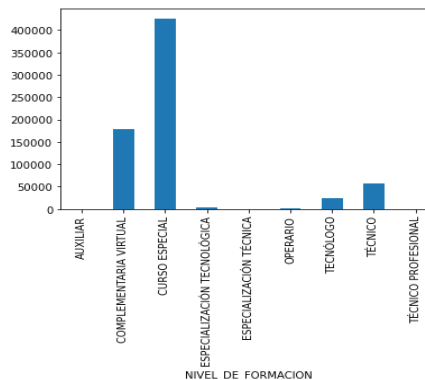
	Frecuencia
SISTEMAS	17363
ENGLISH DOT WORKS 1 (INGLES)	13520
AVANZADO TRABAJO SEGURO EN ALTURAS	11649
ADMINISTRACION DE RECURSOS HUMANOS	8421
CONCEPTUALIZACION DEL LENGUAJE DE PROGRAMACION C++	8320
...	...
AUTOMATIZACION DE SISTEMAS INDUSTRIALES CON PLC	7
PRINCIPIOS BASICOS DE LOS SISTEMAS DOMOTICOS E INMOTICOS	7
TEJIDOS DE CADENERIA EN JOYERIA.	6
PROMOCION EN PROCESOS DE PARTICIPACION Y GESTION AMBIENTAL COMUNITARIA	6
PRODUCCION DE CONTENIDOS ANIMADOS PARA LA SIMULACION DE PROCESOS INDUSTRIALES	1

2397 rows x 1 columns

Tabla 2. Tabla de Frecuencia Absoluta por Programa.

El análisis exploratorio muestra que de acuerdo con los niveles de formación los de mayor interés son

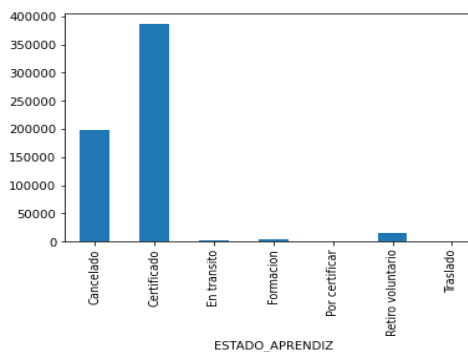
cursos especiales (cursos complementarios presenciales) y complementaria virtual, como lo refleja la gráfica 1.



Gráfica 1. Tendencia Nivel de Formación.

Como resultado del análisis de los datos y teniendo en cuenta que los aprendices pueden matricularse solamente en un curso de formación técnica, tecnológica o especialización tecnológica y estos programas no representan la mayoría de las matrículas, los datos tomados para esta propuesta son de los cursos complementarios virtuales y presenciales ofertados por el Centro de Teleinformática y producción industrial de la regional Cauca.

De igual manera los datos analizados nos muestran el comportamiento del estado del aprendiz, por un lado, la gran mayoría de los usuarios se certifican, pero por otro el porcentaje de cancelación en la formación es alto, como se muestra en la gráfica 2.



Gráfica 2. Estado del Aprendiz.

## 2.3. Preparación de los Datos.

Con la utilización de técnicas de minería de datos se procede a: selección, limpieza, creación de variables necesarias, cambio de formato si es necesario, eliminación de datos innecesarios para así crear el

dataset con el que se desarrollará el modelo. Para seleccionar los datos se toma como subconjunto los cursos de formación complementaria y se eliminan variables que no son necesarias para el proceso. El resultado un dataset con variables como lo muestra la tabla 3.

#	Column	Non-Null Count	Dtype
0	CODIGO_PROGRAMA	605241 non-null	int64
1	PROGRAMA	605241 non-null	object
2	TIPO_DOCUMENTO	605241 non-null	object
3	NUMERO_DOCUMENTO	605241 non-null	object
4	NOMBRE	605241 non-null	object
5	PRIMER_APELLIDO	605241 non-null	object
6	SEGUNDO_APELLIDO	576186 non-null	object
7	ESTADO_APRENDIZ	605241 non-null	object

dtypes: int64(1), object(7)

Tabla 3. Dataset Formación Complementaria.

Una vez realizado el proceso anterior da como resultado la depuración de los datos y así dar mayor precisión, en este caso los programas técnicos, tecnólogos o especializaciones tecnológicas ya no se encuentran en el dataset, ejemplo, el programa técnico de Sistemas, el de mayor demanda, al no ser viable para esta propuesta se elimina. Los nuevos datos se representan mediante una tabla de frecuencia con los cursos de formación complementaria más solicitados por los usuarios como lo muestra la tabla 4.

	Frecuencia
ENGLISH DOT WORKS 1 (INGLES)	13520
AVANZADO TRABAJO SEGURO EN ALTURAS	11649
ADMINISTRACION DE RECURSOS HUMANOS	8421
CONCEPTUALIZACION DEL LENGUAJE DE PROGRAMACION C++	8320
MANEJO BÁSICO DE HERRAMIENTAS OFIMATICAS I	8011
...	...
AUTOMATIZACION DE SISTEMAS INDUSTRIALES CON PLC	7
PRINCIPIOS BASICOS DE LOS SISTEMAS DOMOTICOS E INMOTICOS	7
DISEÑO GRÁFICO APLICADO EN COREL DRAW X5	7
TEJIDOS DE CADENERIA EN JOYERIA.	6
PROMOCIÓN EN PROCESOS DE PARTICIPACIÓN Y GESTIÓN AMBIENTAL COMUNITARIA	6

2176 rows x 1 columns

Tabla 4. Tabla de Frecuencias Formación Complementaria.

El estado del aprendiz es la variable que se utiliza para determinar en cada usuario la valoración de los cursos que ha matriculado, en esta variable encontramos: certificado, por certificar, formación, retiro voluntario, cancelado, en tránsito y traslado. como se muestra a continuación en la tabla 5.

Certificado	386362
Cancelado	197607
Retiro voluntario	15042
Formacion	4390
En transito	1590
Por certificar	191
Traslado	59

Name: ESTADO\_APRENDIZ, dtype: int64

Tabla 5. Estados del Aprendiz.

Es necesario depurar del dataset los registros con los estados: **En Tránsito**, cuando el curso no inicia al no cumplir con el mínimo de aprendices requeridos o no hay disponibilidad de ambientes. **Traslado**, cuando el aprendiz se cambia de sede o de horario.

De la misma forma se elimina registros con datos duplicados, datos nulos, cursos que son obligatorios o no dan valor al perfil de los usuarios como, por ejemplo, curso de inducción o reinducción, curso de inducción a la formación entre otros, y por último se reinicia los índices desde cero dando como resultado una depuración como se muestran en la tabla 6.

Certificado	322838
Cancelado	165946
Retiro voluntario	12700
Formacion	3680
Por certificar	153

Name: ESTADO\_APRENDIZ, dtype: int64

Tabla 6. Depuración del Estado del Aprendiz.

Del nuevo dataset de formación complementaria se extraen subconjuntos de datos que corresponde a los datos de los aprendices (usuarios), los datos de los cursos (ítems) y un dataset que consolida el número de documento de los aprendices, el código de los programas de formación complementaria y el estado del aprendiz, en este último dataset se transforma la columna estado del aprendiz en datos numéricos de la siguiente manera: Certificado = 4, Por Certificar = 4, En Formación = 3, Retiro Voluntario = 2, Cancelado = 1, estos datos numéricos representan las valoraciones del curso.

Las valoraciones anteriormente propuestas se toman como las calificaciones de los cursos en el nuevo dataset que se utiliza para entrenar el modelo del sistema recomendador y su estructura se muestra en la tabla 7.

	NUMERO_DOCUMENTO	CODIGO_PROGRAMA	ESTADO_APRENDIZ
0	1037595113	21450168	2
1	79739677	21450168	4
2	1022375799	21450168	1
3	1081905942	21450168	1
4	80012736	21450168	1

Tabla 7. Dataset para Entrenamiento.

## 2.4. Modelado.

La técnica de recomendación a utilizar para el desarrollo de esta propuesta es el filtrado colaborativo, basada en las valoraciones otorgadas por los usuarios sin depender de la información adicional. Para ello se necesita relacionar dos entidades diferentes:

Los usuarios son los aprendices.

Los ítems son los cursos de formación complementaria.

Es necesario para la construcción del sistema recomendador crear usuarios de entrada con los datos de nombre del programa, estado del aprendiz, y su respectiva valoración como se muestra en la imagen 1.

```

userInput = [
  {'PROGRAMA': 'INGLÉS I', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'FUNDAMENTOS E INFRAESTRUCTURA DE SEGURIDAD EN REDES DE COMPUTADORES', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'VIRTUALIZACIÓN DE SERVIDORES', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'UTILIZACIÓN DE HERRAMIENTAS TIC EN LA ORIENTACIÓN DE PROCESOS FORMATIVOS', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'MARCO NORMATIVO DEL DEPORTE, LA RECREACIÓN Y LA ACTIVIDAD FÍSICA EN COLOMBIA', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'COMPROMISO CIUDADANO ANTE EL ESPECTÁCULO DEL FÚTBOL', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'ELABORACIÓN DE GUÍAS DE APRENDIZAJE', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'ESPECIFICACIÓN DE REQUISITOS DE SOFTWARE', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'EJERCICIO DERECHOS FUNDAMENTALES EN EL TRABAJO', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'DESARROLLO WEB CON PHP', 'ESTADO_APRENDIZ': 1},
  {'PROGRAMA': 'DESARROLLO DE HABILIDADES PARA EL DESEMPEÑO DEL INSTRUCTOR SENIA', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'HERRAMIENTAS PARA LA PLANEACIÓN ESTRATÉGICA', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'ASESORÍA PARA EL USO DE LAS TIC EN LA FORMACIÓN', 'ESTADO_APRENDIZ': 4},
  {'PROGRAMA': 'ORIENTACIONES PEDAGÓGICAS DE LA FORMACIÓN PROFESIONAL BAJO EL ENFOQUE DE COMPETENCIAS', 'ESTADO_APRENDIZ': 4}
]

```

Imagen 1. Creación de usuario.

El algoritmo emplea los datos de calificaciones de los usuarios para predecir, técnicas estadísticas encuentran el conjunto de ítems conocidos como vecinos. Esta vecindad se establece por la similitud que existe entre ellos [5]. Para esto se encuentra en el dataset el código de los programas de formación y su valoración de acuerdo con el nuevo usuario, como se muestra en la tabla 8.

	CODIGO_PROGRAMA	PROGRAMA	ESTADO_APRENDIZ
0	41430001	ASESORIA PARA EL USO DE LAS TIC EN LA FORMACION	4
1	21730185	DESARROLLO WEB CON PHP	1
2	41310120	INGLÉS I	4
3	21450062	FUNDAMENTOS E INFRAESTRUCTURA DE SEGURIDAD EN ...	4
4	21720166	VIRTUALIZACION DE SERVIDORES	4
5	41311369	ESPECIFICACION DE REQUISITOS DE SOFTWARE	4
6	41310304	INGLÉS I	4
7	41311176	UTILIZACION DE HERRAMIENTAS TIC EN LA ORIENTAC...	4
8	42110153	EJERCICIO DERECHOS FUNDAMENTALES EN EL TRABAJO	4
9	41440006	ORIENTACIONES PEDAGOGICAS DE LA FORMACIÓN PROF...	4

Tabla 8. Programa de formación para el usuario.

Posteriormente se agrupa aquellos aprendices que hayan cursado los mismos programas de formación que el usuario creado, el resultado se muestra en la tabla 9.

	NUMERO_DOCUMENTO	CODIGO_PROGRAMA	ESTADO_APRENDIZ
186	1057574103	41430001	1
192	1098799088	21730185	1
198	1110513054	21730185	1
220	38656951	41430001	4
235	1022328773	41430001	1
...	...	...	...
589099	1002968664	42110153	4
589148	1004621158	42110153	4
589465	1061770755	42110153	4
589497	1061225888	42110153	4
589725	1002948352	42110153	4

12048 rows × 3 columns

Tabla 9. Agrupación Similares al usuario.

Para poder resolver la similitud entre dos ítems se utiliza la medida de coeficiente de correlación de Pearson, derivado de las fórmulas de regresión lineal; La relación entre elementos es lineal, los errores independientes y la distribución tiene varianza constante y media 0. El peso asignado al usuario  $u$  para predecir al usuario activo  $a$  dado por ( $r_{a,i}$  es la votación del usuario  $a$  al elemento  $i$ ). [6]

$$w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sigma_a \sigma_u}$$

De tal manera que para que el sistema pueda realizar la recomendación tomará las calificaciones dadas por cada vecino y proporcionarle a su correlación con el usuario actual.

$$p_{a,i} = r_a + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

Ahora para mejorar la predicción es introducir variables que dependan de la calificación es más alta que otras y para proporcionarlos se debe normalizar respecto a la media de la calificación de cada usuario. [6].

$$p_{a,i} = r_a + \sigma_a \frac{\sum_{u=1}^n \frac{(r_{u,i} - \bar{r}_u)}{\sigma_u} * w_{a,u}}{\sum_{u=1}^n w_{a,u}}$$

Se agrupan los aprendices por número de documento y se ordenan dando prioridad a aquellos con mayor número de cursos realizados, como lo muestra la tabla 10.

	NUMERO_DOCUMENTO	CODIGO_PROGRAMA	ESTADO_APRENDIZ
19743	13069292	21730185	1
19744	13069292	41430001	4
317449	13069292	41310120	4
349505	13069292	21450062	4
349506	13069292	21720166	4
349507	13069292	41311176	4
349508	13069292	41311292	4
349509	13069292	66420045	4
349510	13069292	41310603	4
349512	13069292	41311369	4
349513	13069292	42110153	4
349514	13069292	41311584	4
349515	13069292	41311230	4
349516	13069292	41440006	4
100599	1024547016	21730185	1

Tabla 10. Agrupación de aprendices.



Para el Calculo de Recall se tiene en cuenta la pregunta de la encuesta ¿De los cursos que usted conoce y se imparten en el SENA cuál de los que usted no encuentra en la anterior lista, le gustaría cursar? Como se muestra en el ejemplo de la tabla 13.

2. De los cursos que usted conoce y se imparten en el SENA cuál de los que usted no encuentra en la anterior lista, le gustaría cursar.

1.	
2.	
3.	

Tabla 13. Pregunta a usuarios.

Como resultado de la evaluación aplicando las fórmulas de Precisión y Recall se obtiene como lo muestra tabla 14.

No Enc.	Precisión	Recall
1	1.00	0.71
2	0.83	0.83
3	0.67	1.00
4	0.83	1.00
5	1.00	1.00
6	0.83	1.00
7	0.17	0.50
8	0.33	1.00
9	0.67	0.80
10	1.00	0.86
11	0.83	1.00
12	0.83	0.62
13	0.50	0.50
14	0.67	1.00
15	1.00	1.00
16	0.83	1.00
17	0.83	0.83
18	0.50	0.50
19	0.83	1.00
20	1.00	0.50

Tabla 14. Resultado Precisión y Recall.

Se puede concluir que la eficiencia del sistema de recomendación como lo muestra la tabla 15.

Precisión	Recall
76%	83%

Tabla 15. Resultado de Eficiencia.

### 3. RESULTADOS.

Con el propósito de comprobar la exactitud de los resultados obtenidos entre el usuario ingresado y los usuarios de similar comportamiento obtenidos en la matriz de similitudes [7], mostrando los aprendices que presentan similitud, como lo muestra la tabla 16.

similarityIndex	NUMERO_DOCUMENTO	
0	1.0	13069292
70	1.0	1061715358
132	1.0	10296705
137	1.0	1061695281
140	1.0	10002378
146	1.0	1005101320
155	1.0	1014245122
168	1.0	10294917
189	1.0	1049412207
201	1.0	10542346
213	1.0	1061687559

Tabla 16. Usuarios Similares.

Con base a la lista de similitudes entre usuario y cursos de formación complementaria se muestra un listado de cursos posibles recomendaciones para el usuario ingresado con su valoración, como se muestra en la tabla 17.

similarityIndex	NUMERO_DOCUMENTO	CODIGO_PROGRAMA	ESTADO_APRENDIZ	
0	1.0	13069292	21730185	1
1	1.0	13069292	41430001	4
2	1.0	13069292	41310120	4
3	1.0	13069292	21450062	4
4	1.0	13069292	21720166	4
...	...	...	...	...
125	1.0	76311174	41311325	4
126	1.0	76311174	22620440	4
127	1.0	76311174	41311176	4
128	1.0	76311174	41311584	4
129	1.0	76311174	41311584	1

Tabla 17. Posibles cursos para recomendar.

Se muestra un nuevo DataFrame con el promedio ponderado por curso a recomendar, como lo muestra la tabla 18.

recommendation	CODIGO_PROGRAMA	
6110097	1.000000	6110097
11230017	4.000000	11230017
13120325	4.000000	13120325
13410017	4.000000	13410017
13510005	4.000000	13510005
13530036	4.000000	13530036
21450024	4.000000	21450024
21450062	4.000000	21450062
21450168	1.000000	21450168
21450171	2.000000	21450171

Tabla 18. Ponderado de cursos.

Con esto busca los primeros cursos a recomendar identificados con su respectivo código, como se observa en la tabla 19.

CODIGO_PROGRAMA	recommendation	CODIGO_PROGRAMA
93310012	4.0	93310012
21730081	4.0	21730081
13530036	4.0	13530036
21450024	4.0	21450024
21450062	4.0	21450062
93310011	4.0	93310011
93310003	4.0	93310003
93310000	4.0	93310000
41311166	4.0	41311166
21730162	4.0	21730162

Tabla 19. Primeros cursos para recomendar.

Por último, se muestra el nombre de los cursos a recomendar para el usuario ingresado, como lo indica la tabla 20.

CODIGO_PROGRAMA	PROGRAMA
33	41311166 FORMACION TECNOPEdagogICA EN AMBIENTES VIRTUAL...
376	41310121 INGLÉS II
966	41311149 APLICACION DE SOFIAPLUS EN LA ADMINISTRACION E...
1272	41311176 UTILIZACION DE HERRAMIENTAS TIC EN LA ORIENTAC...
1412	33110233 PRIMEROS AUXILIOS
1471	42110159 ORIENTACION DEL PROYECTO DE VIDA CON EL USO DE...
1475	93310012 ELABORACIÓN DE MUEBLES CONTEMPORANEOS
1564	11230017 EVALUADOR DE COMPETENCIAS LABORALES
1679	41311148 BIENVENIDA A INSTRUCTORES "SENA"
2184	51320023 EJECUCIÓN DE INSTRUMENTOS MUSICALES DE CUERDA

Tabla 20. Nombres de cursos a recomendar.

#### 4. DISCUSIÓN (O ANÁLISIS DE RESULTADOS)

De acuerdo con los resultados obtenidos por la ejecución del sistema de recomendación se observa que la mayoría de los cursos recomendados son efectivos y se relacionan con la línea que los usuarios han desarrollado, sin embargo, otros resultados se alejan con respecto a la línea formativa.

La métrica utilizada para medir las predicciones y las recomendaciones obtenidas en su mayoría tiene un valor de 0, lo que puede interpretarse como un resultado negativo, sin embargo, el resultado puede ser debido a la gran cantidad de cursos virtuales y presenciales que ofrece la entidad y la variabilidad de cursos que los usuarios han realizado.

Se busca verificar la confiabilidad del sistema recomendador creando nuevos usuarios con datos reales de los cursos y el estado de cada uno de ellos, de tal manera que permita verificar personalmente

con los usuarios observando si los resultados son cercanos a las recomendaciones que esperaría o por el contrario se aleja de lo esperado; la gran mayoría de los usuarios afirma que la lista de los cursos recomendados es de su interés y respetan la línea de formación.

#### 5. CONCLUSIONES

El artículo propone un sistema recomendador para cursos virtuales y presenciales que ofrece la entidad gubernamental de Colombia "Servicio Nacional de Aprendizaje" – SENA, basado en filtrado colaborativo que utiliza el coeficiente de correlación de Pearson, el cual crea la valoración de los cursos como clasificación a partir del estado en el que terminaron, valorándose de acuerdo con: cancelados, retiro voluntario, en formación, por certificar o certificados, es decir, que los cursos no obtuvieron una calificación directa de quienes los realizaron. Este sistema recomendador se basa en usuarios del centro de teleinformática y producción industrial de la regional Cauca.

Los sistemas recomendadores son utilizados ampliamente en empresas de recreación y ocio, esperando en el futuro que estos tipos de sistemas se utilicen de igual manera en instituciones educativas para facilitar la demanda de estudiantes y sus preferencias.

Para obtener mayor cobertura en el sistema se debe introducir más datos externos en el caso de cursos ofrecidos por otros centros y al tener mucha información el análisis exploratorio, la limpieza y preparación de los datos es indispensable para minimizar fallas en el funcionamiento del sistema recomendador, que cuyo objetivo es generar impacto como método novedoso a nivel institucional tanto regional como nacional.

#### 6. AGRADECIMIENTOS

Los autores expresan su agradecimiento al Centro de Teleinformática y producción Industrial del servicio Nacional de Aprendizaje SENA, en regional Cauca del municipio de Popayán, por permitir acceso y uso de la información de los cursos de formación complementaria, que cuyo objetivo es generar impacto como método novedoso a nivel institucional tanto regional como nacional.

#### 7. REFERENCIAS BIBLIOGRÁFICAS



---

[1] Y. Ocaña Fernandez, L. A. Valenzuela Fernandez, L. L. Garro Aburto, Inteligencia artificial y sus implicaciones en la educación superior. (2019).

[2] X. Cadena Diaz, N, Carabton Galeano,. Planteamiento de Un Sistema Recomendador Basado En Contenido y Colaborativo Para La Elección de Libros - Hdl:11349/7498, n.d.)

[3] Ley N° 119/1994. Servicio Nacional de Aprendizaje - SENA | SITEAL. (n.d.), Disponible en <https://siteal.iiep.unesco.org/bdnp/3161/ley-1191994-servicio-nacional-aprendizaje-SENA>

[4] M. E. Charnelli, L. Lanzarini, J. Diaz, Sistemas Recomendadores aplicados en Educación. (2019).

[5] V. Galan Cortina, Tutora, E. Castro Galán, Aplicación de la Metodología CRISP-DM a un Proyecto de Minería de Datos en el Entorno Universitario. Universidad Carlos III de Madrid. (2015).

[6] S. M. Galan Nieto. Filtrado Colaborativo y Sistemas de Recomendación. (2007).

[7] E. A. Cingolani, *Evaluación de Sistemas Recomendadores de Contenidos Educativos a través de Estudios de Usuarios*, (2014).