



UNIVERSIDAD DE CASTILLA - LA MANCHA

GUÍA DOCENTE

1. DATOS GENERALES

Asignatura: MINERÍA DE DATOS	Código: 42348
Tipología: OPTATIVA	Créditos ECTS: 6
Grado: 406 - GRADO EN INGENIERÍA INFORMÁTICA (AB)_20	Curso académico: 2020-21
Centro: 604 - E.S. DE INGENIERIA INFORMATICA ALBACETE	Grupo(s): 15
Curso: 4	Duración: Primer cuatrimestre
Lengua principal de impartición: Español	Segunda lengua: Inglés
Uso docente de otras lenguas:	English Friendly: N
Página web:	Bilingüe: N

Profesor: JOSE ANTONIO GAMEZ MARTIN - Grupo(s): 15				
Edificio/Despacho	Departamento	Teléfono	Correo electrónico	Horario de tutoría
ESII/1.C.13	SISTEMAS INFORMÁTICOS	2473	jose.gamez@uclm.es	https://www.esiib.uclm.es/pers.php?codpers=56

2. REQUISITOS PREVIOS

Para cursar esta materia es aconsejable haber cursado los módulos de Formación Básica (Módulo I) y el módulo Común a la Rama de Informática (Módulo II).

Esta asignatura se apoya en las competencias y los conocimientos adquiridos en las asignaturas:

- Lógica
- Estadística
- Diseño de Algoritmos
- Sistemas inteligentes
- Sistemas Basados en el Conocimiento

3. JUSTIFICACIÓN EN EL PLAN DE ESTUDIOS, RELACIÓN CON OTRAS ASIGNATURAS Y CON LA PROFESIÓN

La asignatura se ubica en la intensificación de Computación, en el ámbito de las asignaturas de la Inteligencia Artificial. La Minería de datos, y el aprendizaje automático en general, están ligados al campo de la estadística y de la algorítmica, y abordan las técnicas para la extracción de conocimiento implícito en conjuntos de datos. En los últimos años, estas disciplinas están ganando importancia debido al incremento en la producción de datos -propiciado por fenómenos como por ejemplo el auge de internet o las redes sociales - o el desarrollo de nuevas técnicas para la obtención de información genética. Desde el punto de vista profesional, cada vez existe una mayor demanda de analistas de datos en ámbitos tan diversos como el marketing, el análisis de mercados, la seguridad, o la biología.

4. COMPETENCIAS DE LA TITULACIÓN QUE LA ASIGNATURA CONTRIBUYE A ALCANZAR

Competencias propias de la asignatura

Código	Descripción
CM05	Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes o entornos inteligentes.
CM07	Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
INS05	Capacidad para argumentar y justificar lógicamente las decisiones tomadas y las opiniones.
UCLM03	Correcta comunicación oral y escrita.

5. OBJETIVOS O RESULTADOS DE APRENDIZAJE ESPERADOS

Resultados de aprendizaje propios de la asignatura

Descripción

Conocer y desarrollar técnicas de aprendizaje computacional, tanto supervisado como no supervisado, y diseñar e implementar aplicaciones y sistemas que las utilicen.

Desarrollar e implementar un sistema de recuperación de la información de tamaño pequeño a mediano.

Describir y aplicar las distintas fases del proceso de descubrimiento de extracción del conocimiento a partir de grandes volúmenes de datos.

Resultados adicionales

Obtener resultados concluyentes fruto del proceso de extracción del conocimiento y ser capaz de presentarlos y justificarlos

6. TEMARIO

Tema 1: Introducción a la minería de datos

Tema 2: Proceso de descubrimiento de conocimiento a partir de datos.

Tema 3: Validación y evaluación de modelos.

Tema 4: Métodos basados en analogía (kNN)

Tema 5: Predicción numérica: regresión.

Tema 6: Árboles: clasificación y regresión.

Tema 7: Clasificadores probabilísticos

Tema 8: Reducción de la dimensionalidad.

Tema 9: Multclasificadores (ensembles)

Tema 10: Redes neuronales.

Tema 11: Clustering: Técnicas de agrupamiento/segmentación.

Tema 12: Reglas de asociación.

Tema 13: Detección de anomalías.

COMENTARIOS ADICIONALES SOBRE EL TEMARIO

7. ACTIVIDADES O BLOQUES DE ACTIVIDAD Y METODOLOGÍA

Actividad formativa	Metodología	Competencias relacionadas (para títulos anteriores a RD 822/2021)	ECTS	Horas	Ev	Ob	Descripción
Enseñanza presencial (Teoría) [PRESENCIAL]	Método expositivo/Lección magistral	CM05 CM07 INS05	1.26	31.5	N	-	Se usará para introducir los principales conceptos de cada tema.
Prueba final [PRESENCIAL]	Pruebas de evaluación	CM05 CM07 INS05 UCLM03	0.1	2.5	S	S	Corresponde al examen oficial de la asignatura. De carácter individual.
Talleres o seminarios [PRESENCIAL]	Método expositivo/Lección magistral	CM05 CM07 INS05	0.08	2	N	-	Se realizará un seminario en laboratorio de las herramientas/APIs usadas para la implementación de las prácticas e ilustración de los casos de estudio (Python, Weka, etc). Corresponde a la práctica 0.
Enseñanza presencial (Prácticas) [PRESENCIAL]	Método expositivo/Lección magistral	CM05 CM07 INS05	0.06	1.5	N	-	Se usará la primera media hora de la primera sesión correspondiente a cada una de las tres prácticas obligatorias para describir el trabajo a realizar, el material disponible y el entregable a presentar.
Prácticas en aulas de ordenadores [PRESENCIAL]	Trabajo dirigido o tutorizado	CM05 CM07 INS05	0.66	16.5	S	S	Trabajo del alumno con supervisión del profesor en las prácticas de la asignatura.
Resolución de problemas o casos [PRESENCIAL]	Resolución de ejercicios y problemas	CM05 CM07 INS05	0.24	6	N	-	Resolución en clase de problemas y casos de estudio relacionados con los distintos temas estudiados.
Estudio o preparación de pruebas [AUTÓNOMA]	Trabajo autónomo	CM05 CM07 INS05	1.56	39	N	-	Se estima que el alumno dedicará aproximadamente 1.3 horas de estudio autónomo por cada hora de lección magistral recibida. Esto incluye la búsqueda y lectura de material adicional.
Otra actividad no presencial [AUTÓNOMA]	Prácticas	CM05 CM07 INS05	0.84	21	N	-	Horas adicionales a las presenciales para completar las prácticas/proyectos de programación.
Elaboración de memorias de Prácticas [AUTÓNOMA]	Trabajo autónomo	CM05 CM07 INS05 UCLM03	0.42	10.5	S	S	Los alumnos han de plasmar en una memoria el trabajo realizado en la práctica, presentar de forma coherente los resultados y obtener las conclusiones de acuerdo a los objetivos inicialmente fijados. En caso de no superar la práctica en evaluación continua, se podrá recuperar en la entrega correspondiente a la convocatoria ordinaria, si bien, el enunciado podrá ser ligeramente diferente. Las prácticas se diseñan para ser realizadas y defendidas por parejas. No obstante en situaciones particulares y pactadas de antemano con el profesor, podrán realizarse de forma individual.
Elaboración de informes o trabajos [AUTÓNOMA]	Aprendizaje basado en problemas (ABP)	CM05 CM07 INS05 UCLM03	0.78	19.5	S	N	Se facilitarán casos de estudio o ejercicios seleccionados, que el alumno debe completar y que serán evaluados mediante la entrega de tareas o cuestionarios en campus virtual. Este actividad es de carácter individual y lleva asociada la evaluación por pares.
Total:				6	150		
Créditos totales de trabajo presencial: 2.4				Horas totales de trabajo presencial: 60			

Ev: Actividad formativa evaluable

Ob: Actividad formativa de superación obligatoria (Será imprescindible su superación tanto en evaluación continua como no continua)

8. CRITERIOS DE EVALUACIÓN Y VALORACIONES

Sistema de evaluación	Evaluación continua	Evaluación no continua*	Descripción
Resolución de problemas o casos	15.00%	15.00%	(ESC) Este 15% de la nota se dividirá entre los casos de estudio/ejercicios proporcionados, en función de la complejidad de las mismas y de la corrección de los ejercicios entregados. Entrega individual. Esta actividad no es obligatoria ni tiene mínimo. Debe entregarse en las fechas de la evaluación continua.
Elaboración de memorias de prácticas	15.00%	15.00%	(INF) Es obligatoria la entrega de las tres memorias de prácticas. Se abrirá plazo de entrega específico para la evaluación no continua.
Realización de prácticas en laboratorio	15.00%	15.00%	(LAB) Se supervisará la eficiencia y eficacia del código generado para las tres prácticas planteadas. Entrega por grupo de prácticas. Se abrirá plazo de entrega específico para la evaluación no continua.
Otro sistema de evaluación	15.00%	15.00%	(PRES) Cada grupo de prácticas deberá presentar su práctica al profesor y responder a las preguntas realizadas. Entrevista por grupo de prácticas, con preguntas individuales. Se abrirá plazo de entrega específico para la evaluación no continua.
Prueba final	40.00%	40.00%	(ESC) En la convocatoria ordinaria esta prueba tendrá un valor del 40% de la nota final de la asignatura. Entrega individual. Prueba común en evaluación continua y no continua.
Total:	100.00%	100.00%	

* En **Evaluación no continua** se deben definir los porcentajes de evaluación según lo dispuesto en el art. 4 del Reglamento de Evaluación del Estudiante de la UCLM, que establece que debe facilitarse a los estudiantes que no puedan asistir regularmente a las actividades formativas presenciales la superación de la asignatura, teniendo derecho (art. 12.2) a ser calificado globalmente, en 2 convocatorias anuales por asignatura, una ordinaria y otra extraordinaria (evaluándose el 100% de las competencias).

Criterios de evaluación de la convocatoria ordinaria:**Evaluación continua:**

- Las hojas de problemas deben entregarse en las fechas asignadas, si llevan asociada evaluación por pares, esta actividad es imprescindible para computar la nota. Esta actividad no es obligatoria, por lo que puede superarse la asignatura sin realizarla.
- Las prácticas deben ser entregadas y defendidas en las fechas asignadas a cada práctica. Se requiere entregar las tres prácticas y superarlas de forma individual (nota ≥ 5)
- La prueba de teoría se realizará en la fecha de la convocatoria ordinaria asignada por el centro. Será necesario superarla con nota ≥ 5 .
- La calificación de la asignatura es:
 $0.4 \cdot \text{teoría} + 0.45 \cdot \text{prácticas} + 0.15 \cdot \text{problemas}$,
 siempre que los mínimos de teoría y prácticas se cumplan, en caso contrario será:
 mínimo(4.0, nota-teoría) si el examen de teoría se realiza o No Presentado si no se hace el examen de teoría.

Originalidad: La entrega de cualquier ejercicio (examen, memoria de prácticas, código, problemas, etc.) supone implícitamente la declaración de originalidad por parte de los autores, por lo que en caso de detección de plagio, copia, etc. se iniciarán las medidas disciplinarias oportunas.

Evaluación no continua:

El esquema de evaluación es el mismo que en la evaluación continua.

Para las prácticas que se haya superado el mínimo en evaluación continua y para los problemas (que no tienen mínimo) se tomará directamente la nota de las entregas ya evaluadas en la evaluación continua. Para las prácticas no superadas en evaluación continua, se abrirá un plazo de entrega específico para esta convocatoria y posteriormente se anunciarán los horarios para su defensa. El enunciado de las prácticas puede ser ligeramente diferente al de la evaluación continua.

- Todas las prácticas deben ser entregadas y superarlas de forma individual (nota ≥ 5)
- La prueba de teoría se realizará en la fecha de la convocatoria ordinaria asignada por el centro. Será necesario superarla con nota ≥ 5 .
- La calificación de la asignatura es:
 $0.4 \cdot \text{teoría} + 0.45 \cdot \text{prácticas} + 0.15 \cdot \text{problemas}$,
 siempre que los mínimos de teoría y prácticas se cumplan, en caso contrario será:
 mínimo(4.0, nota-teoría) si el examen de teoría se realiza o No Presentado si no se hace el examen de teoría.

Originalidad: La entrega de cualquier ejercicio (examen, memoria de prácticas, código, problemas, etc.) supone implícitamente la declaración de originalidad por parte de los autores, por lo que en caso de detección de plagio, copia, etc. se iniciarán las medidas disciplinarias oportunas.

Particularidades de la convocatoria extraordinaria:

- La nota se obtiene como la media aritmética entre teoría y prácticas una vez superadas ambas partes por separado. En caso contrario la nota será No Presentado si no se ha realizado el examen de teoría y mínimo(4.0, nota-teoría) si el examen si ha sido realizado.

- Se abrirá un plazo específico para la entrega y defensa de las prácticas planteadas para la convocatoria ordinaria (evaluación no continua). Las tres prácticas han de estar aprobadas.

- Se considerará la nota de las actividades superadas en convocatoria ordinaria (examen de teoría y prácticas individuales).

Particularidades de la convocatoria especial de finalización:

- Las prácticas a evaluar corresponden a las del curso anterior. Se abrirá un plazo específico para la entrega y defensa de las mismas. La nota media de las tres prácticas debe ser superior a un 5 y al menos dos de ellas deben estar superadas individualmente (nota >=5).

- La nota se obtiene como la media aritmética ntre teoría y prácticas una vez superadas ambas partes por separado. En caso contrario la nota será No Presentado si no se ha realizado el examen de teoría y mínimo(4.0, nota-teoría) si el examen si ha sido realizado.

9. SECUENCIA DE TRABAJO, CALENDARIO, HITOS IMPORTANTES E INVERSIÓN TEMPORAL**No asignables a temas**

Horas	Suma horas
Prueba final [PRESENCIAL][Pruebas de evaluación]	2.5
Talleres o seminarios [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Enseñanza presencial (Prácticas) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Prácticas en aulas de ordenadores [PRESENCIAL][Trabajo dirigido o tutorizado]	16.5
Resolución de problemas o casos [PRESENCIAL][Resolución de ejercicios y problemas]	7.5
Otra actividad no presencial [AUTÓNOMA][Prácticas]	21
Elaboración de memorias de Prácticas [AUTÓNOMA][Trabajo autónomo]	9
Elaboración de informes o trabajos [AUTÓNOMA][Aprendizaje basado en problemas (ABP)]	19

Comentarios generales sobre la planificación: Esta planificación es ORIENTATIVA, pudiendo variar a lo largo del periodo lectivo en función de las necesidades docentes, festividades, o por cualquier otra causa imprevista. La planificación semanal de la asignatura podrá encontrarse de forma detallada y actualizada en la plataforma Campus Virtual (Moodle). Las actividades presenciales se organizan en tres clases de 1.5 horas por semana. Las clases concretas a usar para cubrir los 6 créditos (60 horas presenciales) se anunciarán oportunamente en CampusVirtual.

Tema 1 (de 13): Introducción a la minería de datos

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	1.3
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020

Comentario: Se introducirá en que consiste al minería de datos y sus principales aplicaciones.

Tema 2 (de 13): Proceso de descubrimiento de conocimiento a partir de datos.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020

Comentario: Se describirá el proceso completo de extracción de conocimiento a partir de bases de datos. Se comentará la metodología CRISP-DM, en particular algunas tareas de preprocesamiento como la discretización o la imputación de valores perdidos.

Tema 3 (de 13): Validación y evaluación de modelos.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020

Comentario: Se introducirán las principales técnicas de validación y medidas de error. Se usarán clasificadores ya estudiados para ilustrar los conceptos. Se introducirán problemas particulares para ilustrar la necesidad de otras medidas de evaluación, p.e. problemas con desequilibrio entre clases para introducir el análisis ROC.

Tema 4 (de 13): Métodos basados en analogía (kNN)

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020

Comentario: En este tema se estudiará el método de los vecinos más cercanos. Se verán las aplicaciones a clasificación y regresión.

Tema 5 (de 13): Predicción numérica: regresión.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020

Comentario: Introducción a la predicción numérica. Técnicas basadas en análisis de regresión. Método de optimización basado en gradiente.

Tema 6 (de 13): Árboles: clasificación y regresión.

Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2

Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: Se repasarán los árboles de clasificación (id3-c4.5) y se estudiará su adaptación a los árboles de regresión y de modelos de regresión.	
Tema 7 (de 13): Clasificadores probabilísticos	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: En este tema se revisará el modelo de clasificación probabilística Naive Bayes y se estudiará la regresión logística y modelos de tipo semi-naive Bayes como TAN o KDB.	
Tema 8 (de 13): Reducción de la dimensionalidad.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: Se estudiarán técnicas de reducción de la dimensionalidad (variables) mediante la transformación y selección de variables. Estas técnicas son genéricas pero de especial interés para los métodos probabilísticos (naive Bayes, regresión logística, TAN, AODE, etc...).	
Tema 9 (de 13): Multiclasificadores (ensembles)	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: En este tema se verá como el uso de múltiples modelos y su posterior combinación/agregación suele mejorar la eficacia/acierto en el aprendizaje supervisado.	
Tema 10 (de 13): Redes neuronales.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: Explicaremos un modelo de carácter bio-inspirado con aplicación a clasificación, regresión y con ciertas modificaciones a clustering.	
Tema 11 (de 13): Clustering: Técnicas de agrupamiento/segmentación.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	3
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: En este tema se estudian los paradigmas de clustering y segmentación, sin duda los de mayor uso en aprendizaje no supervisado. Se trata de extraer patrones entre los registros o instancias del conjunto de datos.	
Tema 12 (de 13): Reglas de asociación.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	3.9
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: En este tema se estudia uno de los paradigmas considerados en principio como de minería de datos: el descubrimiento de reglas de asociación. Ahora se trata de aprender relaciones (reglas) entre las variables a partir de amplias colecciones de transacciones.	
Tema 13 (de 13): Detección de anomalías.	
Actividades formativas	Horas
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	2
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	2.6
Grupo 15:	
Inicio del tema: 21-09-2020	Fin del tema: 22-12-2020
Comentario: En este tema se aborda como detectar anomalías o patrones extraños mediante el aprendizaje de modelos que representan la mayoría de los patrones, y clasificando como anomalías aquellos datos que pertenecen con escasa probabilidad al modelo aprendido.	
Actividad global	
Actividades formativas	Suma horas
Prueba final [PRESENCIAL][Pruebas de evaluación]	2.5
Talleres o seminarios [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Enseñanza presencial (Prácticas) [PRESENCIAL][Método expositivo/Lección magistral]	1.5
Prácticas en aulas de ordenadores [PRESENCIAL][Trabajo dirigido o tutorizado]	16.5
Resolución de problemas o casos [PRESENCIAL][Resolución de ejercicios y problemas]	7.5
Estudio o preparación de pruebas [AUTÓNOMA][Trabajo autónomo]	39
Otra actividad no presencial [AUTÓNOMA][Prácticas]	21
Elaboración de memorias de Prácticas [AUTÓNOMA][Trabajo autónomo]	9
Elaboración de informes o trabajos [AUTÓNOMA][Aprendizaje basado en problemas (ABP)]	19
Enseñanza presencial (Teoría) [PRESENCIAL][Método expositivo/Lección magistral]	31.5
Total horas: 149	

10. BIBLIOGRAFÍA, RECURSOS

Autor/es	Título/Enlace Web	Editorial	Población	ISBN	Año	Descripción
Xindong Wu, Vipin Kumar	The Top Ten Algorithms in Data Mining	Chapman and Hall/CRC		9781420089646	2009	
	Manuales de Python.					
Basilio Sierra	APRENDIZAJE AUTOMÁTICO: CONCEPTOS BÁSICOS Y AVANZADOS: ASPECTOS PRÁCTICOS UTILIZANDO EL SOFTWARE WEKA	Prentice-Hall		9788483223185	2006	
García, Salvador, Luengo, Julián, Herrera, Francisco	Data Preprocessing in Data Mining	Springer		978-3-319-10246-7	2015	
José Hernández Orallo, M.José Ramírez Quintana, Cèsar Ferri Ramírez	INTRODUCCIÓN A LA MINERÍA DE DATOS	Pearson		84 205 4091 9	2004	
Pang-Ning Tan, Michael Steinbach, and Vipin Kumar	Introduction to Data Mining	Addison-Wesley Longman Publishing Co		0321321367	2005	
Witten, Frank & Hall	Data Mining: Practical Machine Learning Tools and Techniques	Morgan & Kauffmann		978-0-12-374856-0	2011	
Joel Grus	Data Science from Scratch: First Principles with Python	O'Reilly UK Ltd		978-1492041139	2019	