รายละเอียดของกระบวนวิชา 229351 ปีการศึกษา 2/2563

1. รหัสกระบวนวิชา 229351 **ชื่อกระบวนวิชา การเรียนรู้เชิงสถิติสำหรับวิทยาการข้อมูล (Statistical Learning for Data Science 1)**
2. ผู้สอน อาจารย์ ดร. ดลภาค พรนพรัตน์ **ผู้ช่วยศาสตราจารย์ ดร. พิมผกา ธานินพงศ์**
3. เวลา Sec 01 : Lecture: Tu 14:30-16:30 น. ที่ SCB4202 Lab: F 14:30-16:30 น. ที่ STB107 Sec 02 : Lecture M 11:00-13:00 น. ที่ SCB4202 Lab: Th 11:00-13:00 น. ที่ STB207
4. หน่วยกิต 3 (3-0-6)
5. Website: donlapark.github.io/ds351

แผนการสอน

สัปดาห์ที่	หัวข้อ/รายละเอียด	จำนวนชั่วโมง			กิจกรรมการเรียนการสอน
		บรรยาย	ปฏิบัติ	ฝึกปฏิบัติ	
1	ชี้แจงและแนะนำกระบวนวิชา แนวคิดพื้นฐานของวิทยาการข้อมูล - ทบทวนหลักการเขียนภาษา python	2	2	0	การบรรยายโดยใช้ภาพนิ่งและคอมพิวเตอร์
2-3	พีชคณิตเชิงเส้นสำหรับวิทยาการข้อมูล (Linear algebra) – นิยามและคุณสมบัติของเวกเตอร์และเมทริกซ์ – การคำนวณด้วยเวกเตอร์และเมทริกซ์ – การแปลงเวกเตอร์ด้วยเมทริกซ์ – การแยกส่วนประกอบของเมทริกซ์	2	2	0	- การบรรยายโดยใช้ภาพนิ่งและคอมพิวเตอร์ - ปฏิบัติการคอมพิวเตอร์ด้วยภาษา python
4	การวิเคราะห์องค์ประกอบหลัก (Principal component analysis)	4	4	0	
5	– นิยามของการเรียนรู้เชิงสถิติ – การสร้าง Training set และ Test set – Bias-Variance tradeoff, no-free-lunch theorem – การวัดประสิทธิภาพของโมเดล	2	2	0	
6-8	การวิเคราะห์การถดถอย (Linear regression) – การเตรียมข้อมูลเพื่อการวิเคราะห์การถดถอย ➢ การสร้างตัวแปรและตัวแปรหุ่น ➢ การสร้างเมทริกซ์แผนแบบ – การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย – การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ – การทดสอบความสัมพันธ์ของตัวแปร – การเลือกตัวแปร – การถดถอยแบบริดจ์และแบบ Lasso	8	8	0	
9-12	การวิเคราะห์อนุกรมเวลาสำหรับวิทยาการข้อมูล – การเตรียมข้อมูลเพื่อการวิเคราะห์อนุกรมเวลา – การวิเคราะห์อนุกรมเวลาแบบคลาสสิค ➢ ความเป็นฤดูกาลและความคงที่ ➢ สหสัมพันธ์ในตัวและสหสัมพันธ์ในตัวบางส่วน ➢ เทคนิคการปรับให้เรียบ ตัวแบบการถดถอยในตัว	8	8	0	

	ตัวแบบการเฉลี่ยเคลื่อนที่ ตัวแบบการเฉลี่ยเคลื่อนที่ถดถอยในตัว				
13-15	การวิเคราะห์การถดถอยโลจิสติกส์สำหรับวิทยาการข้อมูล – การเตรียมข้อมูลเพื่อการวิเคราะห์การถดถอยโลจิสติกส์ทวิ – การวิเคราะห์การถดถอยโลจิสติกส์ทวิ ➢ วิธีภาวะน่าจะเป็นสูงสุด – การวิเคราะห์การถดถอยโลจิสติกส์พหุกลุ่ม – การวิเคราะห์การถดถอยโลจิสติกส์อันดับ	8	8	0	

การประเมินผลการเรียนรู้

2.1 สรุปสัดส่วนการให้คะแนน

- **การบ้าน4 ครั้ง คะแนน 10%**
- **ปฏิบัติการ 12 ครั้ง คะแนน 15%**
- **Kaggle คะแนน 15%**
- **สอบกลางภาค คะแนน 30%** (TBD)
- **สอบปลายภาค คะแนน 30% (สอบวันอาทิตย์ ที่ 28 มีนาคม 2564 เวลา 15:30-18:30 น.)**

2.2 การประเมินผลการเรียนการสอน

- นักศึกษาต้องเข้าสอบทั้งกลางภาคและปลายภาค หากนักศึกษาขาดสอบอย่างใดอย่างหนึ่ง นักศึกษาจะได้รับลำดับขั้น F
- วิธีการตัดเกรดแบบอิงกลุ่ม

การบ้าน

- วิชานี้มีการบ้านทั้งหมด 4 ชิ้น เพื่อให้นักศึกษาได้ทบทวนเนื้อหาที่เรียนไปและมีประสบการณ์ในการเขียนโค้ด ขอให้นักศึกษาส่งการบ้านเป็นไฟล์ *.pdf ผ่านทาง Microsoft Teams ภายในเวลาที่กำหนด โดยไปที่กลุ่มของกระบวนวิชา 229351 แล้วเลือกที่แท็บ **งานที่มอบหมาย** หรือ **Assignments**

ปฏิบัติการ (Lab)

- ทุก ๆ วันจันที่ห้อง **STB107 (Sec 01)** หรือ **STB207 (Sec 02)**
- เน้นการวิเคราะห์ข้อมูลด้วยภาษา Python เราจะทำปฏิบัติการผ่าน Google Colab
- ส่งไฟล์ Notebook หลังจากปฏิบัติการทุกครั้งผ่าน Microsoft Teams

Kaggle

- สมัครเป็นสมาชิกเว็บไซต์ kaggle แล้วเข้าไปที่เว็บไซต์ของการแข่งขัน (link อยู่ใน Microsoft Teams)
- สร้างตัวแบบในการทำนายที่ดีที่สุดโดยใช้ training set แล้ว upload การทำนายบน test set ผ่านทางเว็บไซต์ข้างบน
- การแข่งขันสิ้นสุดวันศุกร์ที่ 12 มีนาคม 2564 เวลา 23:59 น.
- เขียนรายงานถึงวิธีที่ใช้และประสิทธิภาพของตัวแบบที่ได้ ความยาว 4-10 หน้า ส่งภายในวันศุกร์ที่ 19 มีนาคม 2564
  สามารถดูตัวอย่างรายงานได้ที่ **donlapark.github.io/ds351**

ทรัพยากรประกอบการเรียนการสอน

1 ตำราและเอกสารหลัก เอกสารประกอบการสอนของอาจารย์
2 เอกสารและข้อมูลแนะนำ 1. Trevor Hastie et al. **An Introduction to Statistical Learning: with Applications in R**. (available online at https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR Seventh Printing.pdf python code of examples in this book: https://github.com/JWarmenhoven/ISLR-python) 2. Rob J Hyndman and George Athanasopoulos, **Forecasting: Principles and Practice**. (available online at https://otexts.com/fpp2) 3. Gilbert Strang, **Linear Algebra and Learning from Data**.