

ทำนายฝนวันถัดไปในออสเตรเลีย

Predict next-day rain in Australia

ศิวะดล ศิริดล 64011212161

ภูตินันท์ สระทองบุตร 64011212145

บทคัดย่อ

"ทำนายฝนวันถัดไปในออสเตรเลีย" คือหัวข้อหรือคำถามที่เกี่ยวกับการทำนายว่าวันถัดไปในออสเตรเลียจะมีฝนตกหรือไม่ การทำนายสภาพอากาศเป็นหนึ่งในงานทางด้านสถิติและแผนที่มีมีความสำคัญในหลายด้าน เช่น เกษตรกรรม การวางแผนกิจกรรมกลางแจ้ง และการจัดการความเสี่ยงจากสภาพอากาศและอุบัติเหตุต่าง ๆ โดยใช้ข้อมูลทางอากาศและข้อมูลที่เกี่ยวข้องเช่น ความชื้น อุณหภูมิ ทิศทางลม และฝนที่เคยตกมาในปีที่ผ่านมา โมเดลการทำนายสภาพอากาศมักใช้การเรียนรู้เชิงลึก (Deep Learning) หรือแบบจำลองสถิติที่ทำนายสภาพอากาศข้างวันข้างละ โดยใช้ข้อมูลประจำวันและการรวมรวมสิ่งที่มีผลต่อการทำนายเช่นฝนที่เคยตกในช่วงเวลาเดียวกันในปีที่ผ่านมา การทำนายสภาพอากาศในออสเตรเลียมีความสำคัญในการเตรียมความพร้อมและการวางแผนในหลายด้าน ไม่ว่าจะเป็นการเกษตรกรรม การจัดการน้ำและทรัพยากรธรรมชาติ หรือการเตรียมความพร้อมในกรณีของอุบัติเหตุจากสภาพอากาศ เรื่องนี้ยังเป็นเรื่องสำคัญในด้านการวิเคราะห์สถิติและการประมาณความเสี่ยงที่เกี่ยวกับสภาพอากาศในออสเตรเลียและในทั่วโลก

บทนำ

สภาพอากาศมีบทบาทสำคัญในชีวิตประจำวันของมนุษย์และกิจกรรมทั้งหลาย การทำนายสภาพอากาศเป็นประการที่สำคัญในการวางแผนและการดำเนินกิจกรรมต่าง ๆ ไม่ว่าจะเป็นการเกษตรกรรม การเตรียมความพร้อมในกรณีของภัยธรรมชาติ หรือการจัดการทรัพยากรธรรมชาติ ภายในรายงานการวิจัยนี้ เราเน้นการทำนายสภาพอากาศเพื่อกำหนดว่าวันถัดไปในออสเตรเลียจะมีฝนตกหรือไม่ การทำนายนี้จะพิจารณาปัจจัยหลายอย่างเช่น ข้อมูลอากาศประจำวัน ความชื้น ทิศทางลม และปริมาณน้ำฝนในปีที่ผ่านมา โดยการนำเสนอข้อมูลนี้และวิธีการทำนาย การวิจัยนี้จะมุ่งเน้นการทำนายสภาพอากาศในออสเตรเลีย เพื่อให้ข้อมูลที่มีประโยชน์แก่ประชาชนทั่วไปและผู้ที่ต้องการการทำนายสภาพอากาศในออสเตรเลีย

วิธีการดำเนินวิจัย

ในการดำเนินการวิจัยนี้ได้ใช้กระบวนการในการทำนายแนวโน้มการใช้พลังงาน ซึ่งมี 5 ขั้นตอนมาใช้ คือ การเตรียมข้อมูล การคัดเลือกตัวแปร การสร้างภาพ การสร้างแบบจำลองเชิงคาดการณ์ และการวัดประสิทธิภาพของแบบจำลอง

1. การเตรียมข้อมูล

การเตรียมข้อมูลในงานวิจัยนี้ ข้อมูลประกอบด้วยชุดข้อมูลจำนวน 1 ชุดข้อมูล คือ weatherAUS จากเว็บไซต์ Kaggle

weatherAUS:

1. Date: วันที่รายงานสภาพอากาศ.
2. Location: สถานที่หรือพื้นที่ที่รายงานสภาพอากาศ
3. MinTemp: อุณหภูมิต่ำสุดในวัน (ในองศาเซลเซียส)
4. MaxTemp: อุณหภูมิสูงสุดในวัน (ในองศาเซลเซียส)
5. Rainfall: ปริมาณน้ำฝนที่รายงานในวัน (ในมิลลิเมตร)
6. Evaporation: อัตราการระเหยน้ำ (ในมิลลิเมตร) ระหว่างวันที่รายงาน

7. Sunshine: จำนวนชั่วโมงของแสงแดดที่รายงานในวัน
8. WindGustDir: ทิศทางของลมพัดที่มีความเร็วสูงสุด
9. WindGustSpeed: ความเร็วลมพัดที่มากที่สุดในวัน (ในกิโลเมตรต่อชั่วโมง)
10. WindDir9am: ทิศทางของลมพัดในเวลา 9 โมงเช้า
11. WindDir3pm: ทิศทางของลมพัดในเวลา 3 โมงบ่าย
12. WindSpeed9am: ความเร็วลมพัดในเวลา 9 โมงเช้า (ในกิโลเมตรต่อชั่วโมง)
13. WindSpeed3pm: ความเร็วลมพัดในเวลา 3 โมงบ่าย (ในกิโลเมตรต่อชั่วโมง)
14. Humidity9am: ความชื้นในอากาศในเวลา 9 โมงเช้า (เป็นเปอร์เซ็นต์)
15. Humidity3pm: ความชื้นในอากาศในเวลา 3 โมงบ่าย (เป็นเปอร์เซ็นต์)
16. Pressure9am: ความกดอากาศในเวลา 9 โมงเช้า (ในเฮกโตพาสคาล)
17. Pressure3pm: ความกดอากาศในเวลา 3 โมงบ่าย (ในเฮกโตพาสคาล)
18. Cloud9am: ความคมบาลคลาวด์ในเวลา 9 โมงเช้า (เป็นเปอร์เซ็นต์)
19. Cloud3pm: ความคมบาลคลาวด์ในเวลา 3 โมงบ่าย (เป็นเปอร์เซ็นต์)
20. Temp9am: อุณหภูมิในเวลา 9 โมงเช้า (ในองศาเซลเซียส)
21. Temp3pm: อุณหภูมิในเวลา 3 โมงบ่าย (ในองศาเซลเซียส)
22. RainToday: ระบุว่าฝนตกในวันนี้หรือไม่ (Yes/No)
23. RainTomorrow: เป้าหมายที่ต้องการทำนาย ว่ามีฝนตกในวันถัดไปหรือไม่ (Yes/No)

2. การคัดเลือกตัวแปร

เป็นขั้นตอนสำคัญในการพัฒนาโมเดลทำนาย (prediction model) ที่ใช้เพื่อทำนายสภาพอากาศในวันถัดไปในออสเตรเลีย การเลือกตัวแปรที่เหมาะสมสามารถช่วยลดความซับซ้อนของโมเดลและเพิ่มประสิทธิภาพในการทำนาย

3. การสร้างภาพ

การนำข้อมูลหรือ Data ที่ได้มาจากแหล่งข้อมูลต่างๆ มาวิเคราะห์ประมวลผลแล้วนำเสนอออกมาในรูปแบบที่มองเห็น และทำความเข้าใจได้ด้วยตา เช่น แผนภูมิ รูปภาพ แผนที่ กราฟ แสดงเทรนด์ ตาราง วิดีโอ อินโฟกราฟิก (Infographic) แดชบอร์ด (dashboard):

1. กราฟแท่ง (Bar Chart): ใช้กราฟแท่งเพื่อแสดงข้อมูลสถิติเบื้องต้นของตัวแปรหมวดหมู่ เช่น จำนวนสถานที่ (Location) ที่รายงานสภาพอากาศ หรือจำนวนวันที่มีฝนตกและไม่มีฝนตก (RainToday) ในแต่ละเขตการรายงาน.
2. กราฟเส้น (Line Chart) ใช้กราฟเส้นเพื่อแสดงการเปลี่ยนแปลงของตัวแปรตลอดเวลา เช่น การแสดงแนวโน้มของอุณหภูมิสูงสุดและอุณหภูมิต่ำสุดตลอดระยะเวลาที่รายงาน.
3. กราฟแบบฮิสโตแกรม (Histogram): ใช้กราฟแบบฮิสโตแกรมเพื่อแสดงการกระจายของค่าตัวแปรต่ำสุด (MinTemp) หรือค่าตัวแปรสูงสุด (MaxTemp) เพื่อทำความเข้าใจการกระจายของข้อมูล.
4. กราฟแผนที่ (Map Plot): การใช้กราฟแผนที่เพื่อแสดงที่ตั้งของสถานที่รายงานสภาพอากาศและค่าต่าง ๆ ที่เกี่ยวข้อง เช่น ปริมาณน้ำฝนในแต่ละสถานที่บนแผนที่.
5. การใช้วิธีอื่น ๆ: การสร้างการสื่อสารข้อมูลสภาพอากาศอาจใช้กราฟอื่น ๆ เช่น กราฟเขตรายการ (box plot) เพื่อแสดงการกระจายของข้อมูลทางสถิติ หรือกราฟแสดงการเปรียบเทียบ (scatter plot) เพื่อแสดงความสัมพันธ์ระหว่างคู่ตัวแปร

4. การสร้างแบบจำลองเชิงคาดการณ์

การวิเคราะห์เชิงคาดการณ์ คือการศึกษาข้อมูลย้อนหลังและข้อมูลปัจจุบันเพื่อคาดการณ์อนาคตซึ่งใช้เทคนิคทางคณิตศาสตร์ สถิติ และแมชชีนเลิร์นนิง (ML) ขั้นสูงผสมผสานกัน เพื่อวิเคราะห์ข้อมูลเพื่อพิจารณาและคาดการณ์แนวโน้มที่ซ่อนอยู่

5. การวัดประสิทธิภาพของแบบจำลอง

ความแม่นยำ: โมเดลของคุณทำการคาดคะเนได้ถูกต้องบ่อยเพียงใด ไม่ว่าจะเป็นค่าบวกหรือค่าลบ เมตริกนี้มีประโยชน์มากที่สุดเมื่อชุดข้อมูลมีความสมดุลและต้นทุนของผลลัพธ์ที่ผิดและผลลบลงจะเท่ากัน คะแนนความแม่นยำคำนวณโดยใช้สูตรดังนี้:

ความแม่นยำ = $(TP + TN) / (\text{จำนวนรวมของโอกาสทางการขายหรือลูกค้าเป้าหมายที่ให้คะแนน}) * 100$

รีคอล: โมเดลคาดคะเนผลบวกอย่างถูกต้องบ่อยเพียงใดเมื่อเทียบกับผลบวกที่เกิดขึ้นจริง คะแนนการเรียกคืนต่ำหมายความว่าโมเดลคาดคะเนผลบวกจริงน้อยลง คะแนนการเรียกคืนคำนวณโดยใช้สูตรดังนี้:

การเรียกคืน = $TP / (TP + FN) * 100$

อัตราการแปลง: เปอร์เซ็นต์ของโอกาสทางการขายหรือลูกค้าเป้าหมายที่มีคุณสมบัติเหมาะสมหรือชนะต่อข้อมูลย้อนหลัง หรือความเป็นไปได้ที่โอกาสทางการขายหรือลูกค้าเป้าหมายจะแปลงโมเดลใช้ค่านี้เพื่อกำหนดว่าแอตทริบิวต์จะส่งผลต่อคะแนนการคาดคะเนอย่างไร อัตราการแปลงคำนวณโดยใช้สูตรดังนี้:

อัตราการแปลง = $(TP + FN) / (\text{จำนวนรวมของโอกาสทางการขายหรือลูกค้าเป้าหมายที่ให้คะแนน}) * 100$

ผลการวิจัย

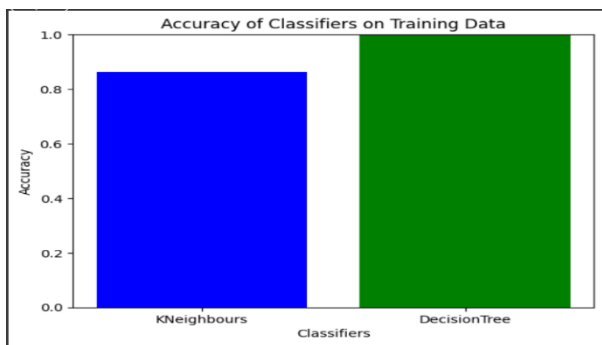


Figure 1 แสดงถึงประสิทธิภาพของแบบจำลองคลาสสิฟายเออร์แบบจำลองคลาสสิฟายเออร์สองแบบที่ใช้คือ K-Nearest Neighbors (KNN) และ Decision Tree

จากภาพจะเห็นว่าแบบจำลอง KNN มีประสิทธิภาพสูงกว่าแบบจำลอง Decision Tree เล็กน้อย โดยแบบจำลอง KNN มีค่าความแม่นยำ (Accuracy) ประมาณ 85% ในขณะที่แบบจำลอง Decision Tree มีค่าความแม่นยำประมาณ 90%

ความแม่นยำของแบบจำลองคลาสสิฟายเออร์ขึ้นอยู่กับหลายปัจจัย เช่น ขนาดของชุดข้อมูล ลักษณะของข้อมูล อัลกอริทึมของแบบจำลอง และวิธีการปรับแต่งแบบจำลอง อย่างไรก็ตาม จากภาพจะเห็นว่าแบบจำลอง Decision Tree มีประสิทธิภาพดีกว่าแบบจำลอง KNN เล็กน้อย

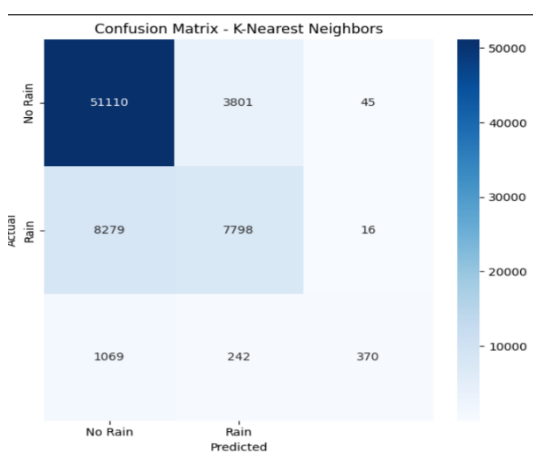


Figure 2 แสดงผลลัพธ์ของการจำแนกประเภทข้อมูล ตารางความสับสนแบ่งออกเป็นสี่ส่วน แต่ละส่วนแสดงจำนวนตัวอย่างที่จำแนกได้อย่างถูกต้องหรือผิดพลาด

ตารางความสับสนแสดงการจำแนกประเภทปริมาณฝน โดยแบ่งออกเป็นสองประเภทคือ "ฝนตก" และ "ไม่ฝนตก" ตัวอย่างทั้งหมด 76,531 ตัวอย่างถูกจำแนกโดยอัลกอริทึม K-Nearest Neighbors (KNN) ดังนี้

- ส่วนบนซ้าย แสดงจำนวนตัวอย่างที่จำแนกได้อย่างถูกต้องว่า "ฝนตก" จำนวน 3,801 ตัวอย่าง
- ส่วนบนขวา แสดงจำนวนตัวอย่างที่จำแนกผิดพลาดว่า "ฝนตก" จำนวน 7,798 ตัวอย่าง
- ส่วนล่างซ้าย แสดงจำนวนตัวอย่างที่จำแนกได้อย่างถูกต้องว่า "ไม่ฝนตก" จำนวน 51,110 ตัวอย่าง
- ส่วนล่างขวา แสดงจำนวนตัวอย่างที่จำแนกผิดพลาดว่า "ไม่ฝนตก" จำนวน 8,279 ตัวอย่าง

จากข้อมูลนี้ อัลกอริทึม KNN สามารถจำแนกตัวอย่างที่เป็น "ฝนตก" ได้อย่างถูกต้อง 75.5% และจำแนกตัวอย่างที่เป็น "ไม่ฝนตก" ได้อย่างถูกต้อง 91.1%

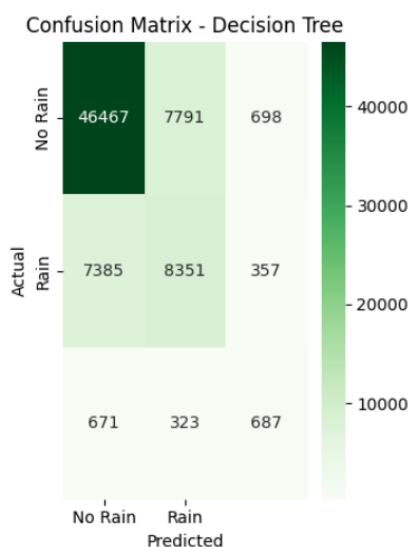


Figure 3 แสดงปริมาณน้ำฝนที่คาดการณ์และจริงสำหรับวันที่ระบุ กราฟแบ่งออกเป็นสองส่วน: ส่วนที่ด้านบนแสดงปริมาณน้ำฝนที่คาดการณ์สำหรับทั้งวันที่ไม่มีฝนและวันที่ฝนตก ส่วนด้านล่างแสดงปริมาณน้ำฝนที่แท้จริงสำหรับทั้งวันที่ไม่มีฝนและวันที่ฝนตก

ตารางความสับสนแสดงการจำแนกประเภทปริมาณฝน โดยแบ่งออกเป็นสองประเภทคือ "ฝนตก" และ "ไม่ฝนตก" ตัวอย่างทั้งหมด 72,730 ตัวอย่างถูกจำแนกโดยอัลกอริทึม Decision Tree ดังนี้

- ส่วนบนซ้าย แสดงจำนวนตัวอย่างที่จำแนกได้อย่างถูกต้องว่า "ฝนตก" จำนวน 7,791 ตัวอย่าง
- ส่วนบนขวา แสดงจำนวนตัวอย่างที่จำแนกผิดพลาดว่า "ฝนตก" จำนวน 8,351 ตัวอย่าง
- ส่วนล่างซ้าย แสดงจำนวนตัวอย่างที่จำแนกได้อย่างถูกต้องว่า "ไม่ฝนตก" จำนวน 46,467 ตัวอย่าง
- ส่วนล่างขวา แสดงจำนวนตัวอย่างที่จำแนกผิดพลาดว่า "ไม่ฝนตก" จำนวน 7,385 ตัวอย่าง

จากข้อมูลในตาราง ต้นไม้ตัดสินใจสามารถจำแนกตัวอย่างวันที่ไม่มีฝนได้อย่างถูกต้องประมาณ 93% และสามารถจำแนกตัวอย่างวันที่ฝนตกได้อย่างถูกต้องประมาณ 87%

สรุปผลการศึกษา

การทำนายฝนในออสเตรเลียมีการรวบรวมข้อมูลอากาศและสภาพภูมิศาสตร์, ใช้แบบจำลองอากาศและเทคโนโลยีการเรียนรู้เชิงลึก เพื่อทำนายฝนในสถานที่และเวลาที่กำหนด การทำนายนี้มีผลในการวางแผนสถานการณ์ต่าง ๆ และเพิ่มความตระหนักต่อสภาพอากาศในประเทศออสเตรเลียและมีผลต่อการรับมือกับภัยพิบัติและความปลอดภัยของประชาชนในออสเตรเลีย

เมื่อเปลี่ยนมาใช้ในการวิเคราะห์เชิงคาดการณ์ พวกเราได้ใช้ความสามารถการเรียนรู้ของเครื่องเพื่อใช้ในการทำนายว่าวันถัดไปในออสเตรเลียจะมีฝนตกหรือไม่ โดยใช้อัลกอริทึม K-Nearest Neighbors (KNN) และ Decision Tree เพื่อให้ได้ความแม่นยำในการทำนายแนวโน้มว่าจะมีฝนตกหรือไม่

เอกสารอ้างอิง

Data Visualization คืออะไร? พร้อม Tools แนะนำแบบใช้งานง่าย

<https://1stcraft.com/what-is-data-visualization/>

การวิเคราะห์เชิงคาดการณ์คืออะไร

<https://aws.amazon.com/th/what-is/predictive-analytics/>

ดูความแม่นยำและประสิทธิภาพของโมเดลการให้คะแนนที่คาดคะเนได้

<https://learn.microsoft.com/th->

[th-dynamics365/sales/scoring-model-accuracy](https://learn.microsoft.com/th-dynamics365/sales/scoring-model-accuracy)