

Application of data mining technology in second-hand car price forecasting

การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลในการพยากรณ์ราคารถยนต์มือสอง

ด้วยการพัฒนาเทคโนโลยีการทำเหมืองข้อมูล ทำให้สามารถจำแนกประเภทหรือคาดการณ์ได้ดีในด้านต่างๆ ประการแรก ได้ดำเนินการสำรวจข้อมูลรถยนต์มือสอง และคุณสมบัติที่มีความสัมพันธ์สูงถึง 99%, พบค่าผิดปกติบางส่วนและค่าที่หายไปโดยดูจาก heat map, box line plot, and violin plot หลังจากลบคุณสมบัติและค่าผิดปกติที่ซ้ำกัน ได้ชุดข้อมูลรถยนต์ที่ใช้แล้วสามชุดโดยใช้วิธีการเดิมสามวิธีสำหรับค่าที่หายไป ได้แก่ การเติมโหมด การเติมค่ามัธยฐาน และ Random forest จากนั้นชุดข้อมูล use d car ทั้งสามชุดจะถูกนำมาใช้เป็นข้อมูลอินพุตเพื่อฝึกโมเดลสี่รุ่น ได้แก่ XGBoost, LightGBM, SVR และโครงข่ายประสาทเทียม BP ได้ทำการทดลองกับชุดทดสอบรถยนต์ใช้แล้ว และผลการทดลองพบว่ารุ่น LightGBM ได้ผลลัพธ์ที่ดีทั้ง MAE และ RA2 โดยให้ผล 0.115350 และ 0.965493 ตามลำดับ ซึ่งดีกว่ารุ่นอื่นๆ อีก 3 รุ่น โครงข่ายประสาทเทียม BP แม้ว่าจะมีประสิทธิภาพที่ดีในการสูญเสียของชุดฝึก อาจเป็นเพราะปัญหาการไล่มากเกินไป ส่งผลให้ชุดทดสอบมีผลลัพธ์ที่ต่ำกว่า ผลลัพธ์ที่ได้จากแบบจำลอง XGBoost นั้นต่ำกว่ารุ่น LightBGM เล็กน้อย เนื่องจากรุ่น LightGBM ได้รับการปรับปรุงในรุ่น XGBoost เพื่อปรับปรุงความเร็วและความแม่นยำในการฝึก โมเดล SVR ให้ผลลัพธ์ที่ดีที่สุดเมื่อเคอร์เนลที่เลือกเป็นแบบเชิงเส้น ซึ่งใช้ไม่ได้กับข้อมูลรถยนต์มือสอง ดังนั้นผลลัพธ์ที่ได้จาก MAE และ RA 2 จึงแยกจากอีกสามรุ่นที่เหลือ

Second-hand Car Price Prediction Based on a Mixed-Weighted Regression Mode

ด้วยการพัฒนายานยนต์ ความต้องการหมุนเวียนของยานยนต์ในรูปแบบของ "รถยนต์มือสอง" ในการเชื่อมโยงการหมุนเวียนเพิ่มขึ้น ในฐานะที่เป็น "สินค้าโภคภัณฑ์ที่อเมริกา" พิเศษ รถยนต์มือสองจึงมีความซับซ้อนมากกว่าสินค้าอเมริกาทั่วไป ด้วยเหตุนี้ จึงเป็นการยากที่จะประเมินราคารถยนต์มือสอง ซึ่งไม่เพียงแต่ได้รับอิทธิพลจากการกำหนดค่าพื้นฐานของรถเท่านั้น แต่ยังรวมถึงสภาพของรถด้วย ปัจจุบันรัฐยังไม่ได้ออกมาตราฐานมาตัดสินมูลค่ารถมือสอง ในการแก้ปัญหา ในบทความนี้ ขั้นแรกให้ทำวิศวกรรมคุณลักษณะ ซึ่งรวมถึงการประมวลผลข้อมูลล่วงหน้าและการคัดกรองคุณลักษณะ การประมวลผลข้อมูลล่วงหน้ารวมถึงการล้างข้อมูลและการแปลงข้อมูล การล้างข้อมูลรวมถึงการลบค่าผิดปกติและการเติมค่าที่ขาดหายไป และการแปลงข้อมูลจะใช้เพื่อรวมรูปแบบข้อมูลเพื่อปรับปรุงคุณภาพข้อมูล การคัดกรองคุณลักษณะประกอบด้วยวิเคราะห์สหสัมพันธ์และการแยกคุณลักษณะตาม LightMBG และคุณลักษณะที่ผ่านการคัดกรองเป็นพื้นฐานสำหรับการสร้างแบบจำลอง การฝึกอบรม และการคาดการณ์ จากนั้น แบบจำลองการถดถอยห้าแบบจะถูกสร้างขึ้นโดยใช้คุณลักษณะคุณลักษณะที่ได้รับจากวิศวกรรมคุณลักษณะสำหรับการฝึกอบรม และประเมินผล จากนั้น Random Forest และ XGBoost จะได้รับการถ่วงน้ำหนักและผสมกันเพื่อให้ได้แบบจำลองการถดถอยแบบใหม่ และผลของแบบจำลองนั้นดีกว่าตัวแบบการถดถอยทั้งห้าแบบ สุดท้ายนี้ แบบจำลองการถดถอยแบบใหม่จะใช้ในการคาดการณ์ราคารถยนต์มือสอง

An Automated Car Price Prediction System Using Effective Machine Learning Techniques