دمج الإحصاءات البايزية مع تحليلات البيانات الضخمة: التحديات والفرص

المدرس المساعد نور علي صعيجل
جامعة البصرة / كلية الادارة والاقتصاد
المدرس المساعد دعاء بشير عباس
جامعة البصرة / كلية التربية الانسانية

المستخلص :

تستكشف هذه الدراسة تطبيق الإحصاءات البايزية لتحليل أسعار السيارات، بهدف تحديد العوامل المهمة التي تؤثر على تقييم السيارة وتطوير نموذج تنبؤي لأسعار السيارات. يستمد البحث دوافعه من تعقيدات سوق السيارات، حيث تعد استراتيجيات التسعير الدقيقة ضرورية للشركات المصنعة والوكلاء. باستخدام مجموعة بيانات شاملة تتضمن سمات مختلفة للسيارة - مثل حجم المحرك ووزن السيارة فارغة وقوة الحصان - يستخدم هذا البحث تقنيات الانحدار البايزي لقياس العلاقات بين هذه الميزات وأسعار السيارات مع دمج عدم اليقين في التوقعات. يكشف التحليل أن وزن السيارة فارغة وحجم المحرك وقوة الحصان ترتبط بشكل إيجابي بأسعار السيارات، في حين أن متوسط استهلاك الوقود في المدينة والطريق السريع لا يؤثر بشكل كبير على التسعير. تُظهر مقاييس تقييم النموذج، بما في ذلك خطأ الجذر التربيعي المتوسط (RMSE) ومتوسط الخطأ المطلق (MAE)، دقة تنبؤية معقولة، مما يؤكد فعالية الأساليب البايزية في هذا السياق. توفر النتائج رؤى قابلة للتنفيذ لأصحاب المصلحة في صناعة السيارات، مما يسلط الضوء على أهمية التركيز على مقاييس الأداء في استراتيجيات التسعير. وفي نهاية المطاف، يساهم هذا البحث في مجموعة المعرفة المتنامية حول التحليلات البايزية وتطبيقاتها العملية في سيناريوهات العالم الحقيقي، مما يشير إلى سبل الاستكشاف المستقبلي في مجال تقييم المركبات.

الكلمات المفتاحية: الإحصاءات البايزية، تحليلات البيانات الضخمة، النمذجة التنبؤية، التحسين البايزي، تقنيات الاستدلال البايزي، مواصفات النموذج، التعاون بين التخصصات.

Integrating Bayesian Statistics with Big Data Analytics: Challenges and Opportunities

Assistant Lecturer Nour Ali Saijil

University of Basra / College of Administration and Economics

Assistant Lecturer Duaa Bashir Abbas

University of Basra / College of Humanities

Abstract :

This study explores the application of Bayesian statistics to analyze car pricing, aiming to identify significant factors influencing vehicle valuation and to develop a predictive model for car prices. The research is motivated by the complexities of the automotive market, where accurate pricing strategies are essential for manufacturers and dealerships. Employing a comprehensive dataset that includes various car attributes—such as engine size, curb weight, and horsepower this research utilizes Bayesian regression techniques to quantify the relationships between these features and car prices while incorporating uncertainty into the predictions. The analysis reveals that curb weight, engine size, and horsepower are positively correlated with car prices, whereas average city and highway fuel economy do not significantly impact pricing. Model evaluation metrics, including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), demonstrate reasonable predictive accuracy, underscoring the effectiveness of Bayesian methods in this context. The findings provide actionable insights for stakeholders in the automotive industry, highlighting the importance of focusing on performance metrics in pricing strategies. Ultimately, this research contributes to the growing body of knowledge on Bayesian analytics and its practical applications in real-world scenarios, suggesting avenues for future exploration within the field of vehicle valuation.

Keyword: Bayesian Statistics, Big Data Analytics, Predictive Modeling, Bayesian Optimization, Bayesian Inference Techniques, Model Specification, Interdisciplinary Collaboration.

المقدمة :

الإحصاءات البايزية هي نهج قوي للاستدلال الإحصائي يستخدم الاحتمالية لقياس عدم اليقين وتحديث المعتقدات في ضوء الأدلة الجديدة. في جوهرها، تفسر الإحصاءات البايزية الاحتمالية كمقياس للثقة أو الاعتقاد فيما يتعلق بوقوع الأحداث، مما يسمح للمحللين بدمج المعرفة السابقة في تحليلاتهم. تستند هذه المنهجية إلى نظرية بايز، التي تصف رياضيًا كيفية تحديث احتمالية فرضية بناءً على البيانات المرصودة. يمكن إرجاع الجذور التاريخية للإحصاءات البايزية إلى عمل توماس بايز في القرن الثامن عشر (Loredo & Wolpert, 2024). مع مساهمات كبيرة من بيير سيمون لابلاس في العقود اللاحقة. على الرغم من مواجهة الشكوك طوال معظم القرن العشرين بسبب التحديات الفلسفية والحسابية، اكتسبت الأساليب البايزية مكانة بارزة في التحليل الإحصائي الحديث، وخاصة مع التقدم في القوة الحسابية والخوارزميات مثل سلسلة ماركوف مونت كارلو (MCMC). اليوم، يتم تطبيق الإحصاءات البايزية على نطاق واسع في مختلف المجالات، بما في ذلك الرعاية الصحية، والتمويل، والتعلم الآلي، مما يوفر إطارًا قويًا لاتخاذ قرارات مستنيرة في ظل عدم اليقين. وتؤكد قدرتها على تحديث المعتقدات الذاتية بشكل عقلاني بناءً على البيانات التجريبية على أهميتها في عمليات تحليل البيانات واتخاذ القرار المعاصرة (Fei et al., 2024).

يشير تحليل البيانات الضخمة إلى عملية فحص مجموعات البيانات الكبيرة والمعقدة للكشف عن الأنماط والارتباطات والرؤى المخفية التي يمكن أن تفيد عملية اتخاذ القرار. تشمل الخصائص الرئيسية للبيانات الضخمة اولها **الحجم** :الكمية الهائلة من البيانات الناتجة عن مصادر مختلفة، بما في ذلك وسائل التواصل الاجتماعي وأجهزة الاستشعار والمعاملات. الخاصية الثانية **السرعة:** السرعة التي يتم بها إنشاء البيانات الجديدة ومعالجتها. ثالثا **التنوع:** الأنواع المختلفة من البيانات (المنظمة وغير المنظمة وشبه المنظمة) التي يمكن تحليلها (Wei et al., 2022). تستخدم تحليلات البيانات الضخمة تقنيات مختلفة، بما في ذلك التعلم الآلي والتحليل الإحصائي واستخراج البيانات، لاستخراج رؤى ذات مغزى من مجموعات البيانات الضخمة. لها تطبيقات في مجالات عديدة مثل التمويل والرعاية الصحية والتسويق والعلوم الاجتماعية, يوفر دمج الإحصاءات البايزية مع تحليلات البيانات الضخمة العديد من الفرص والتحديات تتضمن

الفرص:

- يمكن للأساليب البايزية تعزيز النمذجة التنبؤية من خلال دمج المعرفة السابقة في تحليلات البيانات الضخمة.
- توفر إطارًا لقياس عدم اليقين، وهو أمر ضروري في عمليات صنع القرار القائمة على مجموعات البيانات الضخمة.
- يمكن للأساليب البايزية تحسين قابلية تفسير النموذج والسماح بالنمذجة المرنة للعلاقات المعقدة في البيانات الضخمة (Bharadiya, 2023).

اما التحديات:

- تنشأ التعقيدات الحسابية عند تطبيق الأساليب البايزية التقليدية على مجموعات البيانات الضخمة بسبب الحسابات المكثفة المطلوبة للتوزيعات اللاحقة.
- قد يكون اختيار الأولويات المناسبة أمرًا صعبًا في المساحات عالية الأبعاد النموذجية لسياقات البيانات الضخمة.يمكن أن تؤدي مشكلات جودة البيانات مثل الضوضاء والقيم المفقودة إلى تعقيد التحليل البايزي في إعدادات البيانات الضخمة (Cheng et al., 2022).

باختصار، في حين توفر الإحصاءات البايزية إطارًا قويًا لتحديث المعتقدات استنادًا إلى الأدلة، تركز تحليلات البيانات الضخمة على استخراج الأفكار من كميات كبيرة من البيانات المتنوعة. يقدم تكاملها إمكانيات مثيرة لتطوير التحليل الإحصائي ولكنه يتطلب أيضًا معالجة تحديات حسابية ومنهجية كبيرة.

اهمية البحث

تكمن أهمية هذا البحث في استكشافه للإحصاءات البايزية كأداة حيوية لفهم وتحليل مجموعات البيانات المعقدة، وخاصة في سياق تسعير السيارات. ومع مواجهة صناعة السيارات لمنافسة متزايدة وتغير سريع في تفضيلات المستهلكين، فإن الاستفادة من الأساليب الإحصائية المتقدمة أمر ضروري لاتخاذ قرارات مستنيرة. ومن خلال تطبيق تقنيات الانحدار البايزي على مجموعة بيانات أسعار السيارات، لا يسلط هذا البحث الضوء فقط على العوامل المهمة التي تؤثر على تقييم السيارة - مثل وزن السيارة الفارغ وحجم المحرك وقوة الحصان - بل يوضح أيضًا كيف يمكن للأساليب البايزية قياس عدم اليقين بشكل فعال وتوفير رؤى قوية. يسمح هذا النهج لأصحاب المصلحة بالتنقل بشكل أفضل بين تعقيدات استراتيجيات التسعير وديناميكيات السوق، مما يؤدي في النهاية إلى اتخاذ قرارات أكثر فعالية. وعلاوة على ذلك، تؤكد النتائج على أهمية الإحصاءات البايزية في تحليل البيانات المعاصر، مما يُظهر إمكاناتها لتعزيز الدقة التنبؤية وإبلاغ المبادرات الاستراتيجية عبر مختلف القطاعات. ومع استمرار الصناعات في توليد كميات هائلة من البيانات، يساهم هذا البحث في فهم أعمق لكيفية الاستفادة من المنهجيات البايزية لاستخراج رؤى ذات مغزى وتعزيز الميزة التنافسية.

مشكلة البحث

تدور مشكلة البحث التي تناولتها هذه الدراسة حول تعقيدات تسعير السيارات في سوق السيارات، وخاصة تحدي التنبؤ الدقيق بأسعار المركبات بناءً على عوامل مؤثرة مختلفة. ومع تزايد توافر البيانات حول سمات السيارات - مثل مواصفات المحرك ووزن السيارة الفارغ واقتصاد الوقود - هناك حاجة ملحة لتطوير أساليب تحليلية قوية يمكنها قياس العلاقات بين هذه السمات وأسعار السيارات بشكل فعال. غالبًا ما تفشل الأساليب الإحصائية التقليدية في التقاط حالة عدم اليقين والتباين المتأصلة الموجودة في مثل هذه البيانات، مما يؤدي إلى استنتاجات مضللة محتملة. يسعى هذا البحث إلى سد هذه الفجوة من خلال استخدام الإحصاءات البايزية، والتي تسمح بدمج المعرفة السابقة وتحديث المعتقدات بناءً على أدلة جديدة. من خلال التركيز على تطبيق تقنيات الانحدار البايزي لتحليل مجموعة بيانات أسعار السيارات، تهدف هذه الدراسة إلى توفير فهم أوضح لكيفية تأثير العوامل المختلفة على تقييم السيارة، مما يساهم في نهاية المطاف في اتخاذ قرارات أكثر استنارة للشركات المصنعة والوكلاء والمستهلكين على حد سواء.

اهداف البحث

الأهداف الأساسية لهذا البحث هي دراسة العوامل المؤثرة على أسعار السيارات وتطبيق الأساليب الإحصائية البايزية لتطوير نموذج تنبؤي يعكس هذه العلاقات بدقة. وعلى وجه التحديد، تهدف الدراسة إلى تحديد المتنبئين المهمين لأسعار السيارات من مجموعة بيانات شاملة تتضمن سمات مختلفة للمركبة مثل حجم المحرك ووزن السيارة فارغة وقوة الحصان واقتصاد الوقود. من خلال استخدام تقنيات الانحدار البايزي، يسعى البحث إلى تحديد عدم اليقين المرتبط بتنبؤات الأسعار وتوفير فترات زمنية موثوقة لهذه التقديرات. بالإضافة إلى ذلك، تهدف الدراسة إلى تقييم فعالية الأساليب البايزية مقارنة بالأساليب الإحصائية التقليدية، مع تسليط الضوء على مزاياها من حيث المرونة والقدرة على دمج المعرفة السابقة. في النهاية، يهدف هذا البحث إلى توليد رؤى قابلة للتنفيذ يمكن أن تفيد استراتيجيات التسعير لأصحاب المصلحة في مجال السيارات وتعزيز فهم ديناميكيات السوق في سياق تقييم المركبات.

منهجية البحث

تتضمن منهجية البحث المستخدمة في هذه الدراسة نهجًا منهجيًا لتحليل مجموعة بيانات أسعار السيارات باستخدام تقنيات الانحدار البايزي. في البداية، تم إجراء تحليل استكشافي للبيانات (EDA) لفهم بنية مجموعة البيانات، وتحديد أي قيم مفقودة، وتلخيص الخصائص الرئيسية من خلال الإحصاءات الوصفية. بعد ذلك، تم تحويل المتغيرات الفئوية إلى تنسيقات رقمية باستخدام الترميز الساخن الواحد، وتم تطبيع المتغيرات المستمرة لتعزيز أداء النموذج. ثم تم تحديد نموذج الانحدار البايزي، مع سعر السيارة كمتغير تابع وسمات مختلفة - مثل حجم المحرك ووزن السيارة وقوة الحصان - كمتغيرات مستقلة. استخدم النموذج نظرية بايز لتحديث المعتقدات حول المعلمات بناءً على البيانات المرصودة، باستخدام أساليب سلسلة ماركوف مونت كارلو (MCMC) للعينة اللاحقة. تم تقييم أداء النموذج باستخدام مقاييس مثل خطأ الجذر التربيعي المتوسط (RMSE) ومتوسط الخطأ المطلق (MAE)، والتي قدمت رؤى حول دقة التنبؤ. وأخيرا، تم حساب فترات زمنية موثوقة لقياس عدم اليقين حول توقعات الأسعار، مما يسمح بفهم شامل للعوامل المؤثرة على أسعار السيارات وتداعياتها على أصحاب المصلحة في صناعة السيارات.

حدود البحث

شملت مصطلحات البحث المستخدمة في هذا البحث مجموعة من الكلمات والعبارات الرئيسية ذات الصلة بدراسة تسعير السيارات والإحصاءات البايزية. وشملت المصطلحات الرئيسية "الإحصاءات البايزية" و"الانحدار البايزي" و"الاستدلال البايزي"، والتي كانت ضرورية لفهم التقنيات الإحصائية المطبقة في التحليل. بالإضافة إلى ذلك، تم استخدام مصطلحات مثل "التنبؤ بسعر السيارة" و"العوامل المؤثرة في أسعار السيارات" و"تحليل سوق السيارات" لتركيز البحث على السياق المحدد لتقييم المركبات. وشملت مصطلحات البحث المهمة الأخرى "تحليل البيانات الاستكشافي" و"سلسلة ماركوف مونت كارلو" و"النمذجة التنبؤية"، والتي سلطت الضوء على الأساليب المنهجية المستخدمة في الدراسة. أدى إدراج مصطلحات مثل "معالجة البيانات مسبقًا" و"قياس عدم اليقين" و"الفواصل الزمنية الموثوقة" إلى تحسين البحث بشكل أكبر لالتقاط الفروق الدقيقة للأساليب البايزية وتطبيقها على مجموعات البيانات في العالم الحقيقي. ومن خلال استخدام مصطلحات البحث هذه، هدف البحث إلى جمع رؤى شاملة من الأدبيات والمنهجيات الموجودة، مما يساهم في نهاية المطاف في فهم قوي لكيفية تعزيز التقنيات البايزية لتحليل أسعار السيارات.

هيكلية البحث

يتألف هيكل البحث في هذه الدراسة من عدة فصول رئيسية تتناول بشكل منهجي أهداف ومنهجية تحليل أسعار السيارات باستخدام الإحصاءات البايزية. يضع الفصل التمهيدي الأساس من خلال تعريف الإحصاءات البايزية، وتوفير السياق التاريخي، وتسليط الضوء على أهميتها في التحليل الإحصائي الحديث. بعد **المقدمة**، يقوم **فصل مراجعة الأدبيات** بتلخيص الأبحاث الحالية الأساليب البايزية وتطبيقاتها في مجالات مختلفة، وبالتالي وضع هذه الدراسة ضمن الخطاب الأكاديمي الأوسع. يوضح **فصل المنهجية** النهج المنهجي المتبع في البحث، بما في ذلك جمع البيانات والمعالجة المسبقة وتحليل البيانات الاستكشافي ومواصفات النموذج وتقنيات التقييم المستخدمة لتحليل مجموعة البيانات. **يقدم فصل النتائج** النتائج من تحليل الانحدار البايزي، بما في ذلك المتنبئات المهمة لأسعار السيارات ومقاييس أداء النموذج مثل RMSE وMAE. يتضمن هذا الفصل أيضًا تصورات للتوزيعات اللاحقة والفواصل الزمنية الموثوقة لتوضيح عدم اليقين المرتبط بالتنبؤات. وأخيرًا، يلخص **الفصل الختامي النتائج الرئيسية**، ويناقش آثارها على أصحاب المصلحة في صناعة السيارات، ويقر بحدود الدراسة، **ويقترح سبلًا للبحوث**

المستقبلية. ويضمن هذا النهج المنظم تدفقًا متماسكًا للمعلومات ويسهل الفهم الشامل لكيفية تطبيق الأساليب البايزية لتعزيز تحليل أسعار السيارات.

المرجعات الادبية

الإحصاءات البايزية

الإحصاءات البايزية هي نموذج إحصائي يستخدم نظرية بايز لتحديث احتمالية فرضية ما مع توفر المزيد من الأدلة أو المعلومات. في جوهرها، توفر الإحصاءات البايزية إطارًا لدمج المعتقدات السابقة والبيانات الجديدة لاستخلاص استنتاجات حول المعلمات أو الأحداث غير المعروفة.

تتضمن المكونات الرئيسية للإحصاءات البايزية ما يلي:

1- **نظرية بايز**: تعبر هذه النظرية الأساسية عن العلاقة بين الاحتمالية السابقة والاحتمالية والاحتمالية اللاحقة. يمكن صياغتها رياضيًا على النحو التالي:

$$P\ (A/B) = \left(\frac{P\ (A/B)\ .P(A)}{P(B)}\right)$$

حيث:

P(A|B) هو الاحتمال اللاحق (احتمال الفرضية A بالنظر إلى البياناتB)

P(B|A) هو الاحتمال (احتمال ملاحظة البيانات B بالنظر إلى الفرضيةA)

P(A) هو الاحتمال المسبق (الاعتقاد الأولي حول الفرضيةA).

P(B) هو الاحتمال الهامشي (الاحتمال الإجمالي لملاحظة البياناتB).

- التوزيع المسبق: يمثل هذا الاعتقادات الأولية حول المعلمات قبل ملاحظة أي بيانات. وهو يغلف المعرفة أو الافتراضات الموجودة ويمكن أن يكون ذاتيًا بطبيعته.
- دالة الاحتمال: تقيس مدى احتمالية أن تكون البيانات الملاحظة تحت قيم معلمات مختلفة. وهي تعكس المعلومات التي توفرها البيانات نفسها.
- التوزيع اللاحق: يجمع هذا بين الاعتقادات السابقة واحتمالية أن تنتج البيانات الملاحظة معتقدات محدثة حول المعلمات بعد النظر في أدلة جديدة.

2- **قانون الاحتمال الكلي**

لحساب P(B)، وخاصة في الحالات التي تحتوي على فرضيات متعددة، نستخدم قانون الاحتمال الكلي:

$$P\ (B) = \sum_{i} P(B/A_i)\ P(A_i)$$

تلخص هذه المعادلة كل الفرضيات الممكنة i ، مما يوفر طريقة لحساب الاحتمالية الإجمالية لملاحظة B .

3- **التوزيع الخلفي**

في التحليل البايزي، يجمع التوزيع الخلفي بين المعتقدات السابقة والبيانات الجديدة من خلال نظرية بايز:

$$\boldsymbol{P(\theta/D)} = \left(\frac{\boldsymbol{P\ (D/\theta)\ .\ P(\theta)}}{\boldsymbol{P(D)}}\right)$$

حيث:

$P\ (\theta/D)$ هو التوزيع الخلفي للمعامل θ معطاة البيانات D.

$P\ (D/\theta)$ هو احتمال البيانات المعطاة للمعامل θ

$P(\theta)$ هو التوزيع المسبق للمعامل θ.

والذي يمثل الاحتمالية الهامشية $P(D) = \int P\ P\ (D/\theta)\ .P(\theta)d\theta.$

4- التوزيعات السابقة واللاحقة

يؤثر اختيار التوزيع السابق بشكل كبير على التحليل البايزي. تشمل التوزيعات السابقة الشائعة ما يلي:

التوزيع بيتا (للنسب):

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$$

$$Where\ B\ (\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha,\beta)}$$

التوزيع الطبيعي (للمعلمات المستمرة):

$$P(\mu/\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

يمكن استخلاص التوزيع الخلفي من خلال الجمع بين السابق والاحتمالي:

$$P(\theta/D) \propto P(\ \theta/D)P(\theta)$$

5- التوزيع التنبئي

يسمح التوزيع التنبئي بإجراء تنبؤات حول الملاحظات المستقبلية استنادًا إلى البيانات الموجودة:

$$P(\ y_{new}/D) = \int P(\ y_{new}/\theta)\ P(\theta/D)d(\theta)$$

تتكامل هذه المعادلة مع كل القيم الممكنة لـ θ، مرجحة باحتمالاتها الخلفية.

6- سلسلة ماركوف مونت كارلو (MCMC)

في النماذج المعقدة حيث يصعب الحصول على حلول تحليلية، يتم استخدام طرق MCMC لتقريب التوزيعات اللاحقة:

تتضمن الفكرة الأساسية إنشاء سلسلة ماركوف التي يكون التوزيع اللاحق المطلوب فيها هو توزيع التوازن الخاص بها. تتضمن أكثر طرق MCMC شيوعًا ما يلي:

1. **خوارزمية متروبوليس- هاستينجز:**
   - إنشاء اقتراح لحالة جديدة.
   - قبول أو رفض هذا الاقتراح بناءً على نسبة قبول محسوبة.
2. **أخذ العينات من جيبس:**
   - أخذ عينات متكررة من التوزيعات الشرطية لكل معلمة مع الأخذ في الاعتبار جميع المعلمات الأخرى .

تشكل المعادلات الموضحة في هذا الفصل الأساس للإحصاءات البايزية، حيث توفر أدوات لتحديث المعتقدات في ضوء الأدلة الجديدة وإجراء تنبؤات مستنيرة. يعد فهم هذه المعادلات أمرًا بالغ الأهمية لتطبيق الأساليب البايزية بشكل فعال عبر مجالات مختلفة، بما في ذلك التعلم الآلي والتمويل والرعاية الصحية والمزيد. مع استمرار تطور التقنيات البايزية، فإن تكاملها مع تحليلات البيانات الضخمة سيعزز بشكل أكبر قابلية تطبيقها وتأثيرها في حل المشكلات المعقدة في العالم الحقيقي(Barido-Sottani et al., 2024) .

التطورالتاريخي والأهمية في التحليل الإحصائي

تعود جذور الإحصاءات البايزية إلى عمل توماس بايز، وهو إحصائي ولاهوتي من القرن الثامن عشر، والذي صاغ نظرية بايز التي نُشرت بعد وفاته في عام 1763. اكتسبت الطريقة زخمًا طوال القرن العشرين، وخاصة مع التقدم في التقنيات الحسابية التي مكنت من إنشاء نماذج وتحليلات أكثر تعقيدًا. تكمن أهمية الإحصاءات البايزية في قدرتها على توفير إطار متماسك لاتخاذ القرار في ظل عدم اليقين. على عكس الإحصاءات التكرارية، التي تعتمد غالبًا على معلمات ثابتة وقيمp ، تسمح الأساليب البايزية بتفسير أكثر بديهية للاحتمالية كمقياس للاعتقاد أو الثقة فيما يتعلق بالفرضيات. تجعل هذه المرونة الإحصاءات البايزية ذات قيمة خاصة في مجالات مثل التعلم الآلي والبحث الطبي وتقييم المخاطر، حيث يعد فهم عدم اليقين أمرًا بالغ الأهمية لاتخاذ القرار الفعال (Poole, 2023).

باختصار، تقدم الإحصاءات البايزية أدوات قوية لتحديث المعتقدات بناءً على الأدلة، مما يجعلها نهجًا أساسيًا في التحليل الإحصائي الحديث. ويعكس تطورها التاريخي اعترافًا متزايدًا بأهمية دمج المعرفة السابقة في الاستدلال الإحصائي، مما يمهد الطريق لتطبيقها على نطاق واسع عبر مختلف التخصصات.

تحليلات البيانات الضخمة

يشير تحليل البيانات الضخمة إلى عملية جمع وفحص وتحليل مجموعات البيانات الضخمة والمعقدة للكشف عن الأنماط والارتباطات والرؤى التي يمكن أن تفيد عملية اتخاذ القرار. أصبح هذا المجال بالغ الأهمية حيث تسعى المنظمات إلى الاستفادة من الكميات الهائلة من البيانات التي يتم إنشاؤها في المشهد الرقمي اليوم. تلخيص الخصائص التي تحدد البيانات الضخمة:

- **الحجم:** يشير هذا إلى الكمية الهائلة من البيانات التي يتم إنشاؤها من مصادر مختلفة، بما في ذلك وسائل التواصل الاجتماعي وأجهزة الاستشعار والمعاملات والمزيد. يتم قياس البيانات الضخمة عادةً بالجيجابايت أو التيرابايت أو حتى البيتابايت. يستمر حجم البيانات في النمو بشكل كبير، مما يستلزم طرقًا جديدة للتخزين والتحليل

السرعة: تصف السرعة التي يتم بها إنشاء البيانات ومعالجتها. في العديد من التطبيقات، يتم إنشاء البيانات في الوقت الفعلي أو في وقت قريب من الوقت الفعلي، مما يتطلب معالجة وتحليل سريعين لاستخراج رؤى في الوقت المناسب (Saeejil et al., 2024).

- تؤكد هذه الخاصية على الحاجة إلى أنظمة يمكنها التعامل مع تدفقات البيانات عالية الإنتاجية بكفاءة
- **التنوع:** تأتي البيانات الضخمة في أشكال مختلفة، بما في ذلك البيانات المنظمة وشبه المنظمة وغير المنظمة. يفرض هذا التنوع تحديات من حيث التكامل والتحليل، حيث قد لا تكون قواعد البيانات التقليدية مجهزة للتعامل مع مثل هذه المجموعات غير المتجانسة من البيانات (Wang et al., 2022).

تتضمن الخصائص الإضافية التي ظهرت في المناقشات حول البيانات الضخمة ما يلي:

- **التنوع:** يشير هذا إلى الطبيعة غير المتسقة لتدفقات البيانات. يمكن أن تختلف البيانات بشكل كبير بمرور الوقت أو عبر مصادر مختلفة، مما يعقد جهود التحليل.
- **الصدق:** تتعلق الصدق بجودة ودقة البيانات. يعد ضمان موثوقية البيانات الضخمة أمرًا ضروريًا لاتخاذ قرارات مستنيرة.
- **القيمة:** في النهاية، يتمثل هدف تحليلات البيانات الضخمة في استخراج رؤى قيمة يمكن أن تدفع قرارات واستراتيجيات الأعمال (Batko & Ślęzak, 2022).

يعمل تحليل البيانات الضخمة على تحويل العديد من الصناعات من خلال تمكين المؤسسات من اتخاذ قرارات تعتمد على البيانات. تتضمن بعض الاتجاهات والتطبيقات الحالية ما يلي:

- **الرعاية الصحية:** تُستخدم تحليلات البيانات الضخمة لتحليل سجلات المرضى والتجارب السريرية والنتائج الصحية. من خلال تحديد الأنماط في مجموعات البيانات الضخمة، يمكن لمقدمي الرعاية الصحية تحسين رعاية المرضى وتحسين خطط العلاج وخفض التكاليف
- **التمويل:** في التمويل، تساعد تحليلات البيانات الضخمة المؤسسات على اكتشاف الأنشطة الاحتيالية وتقييم مخاطر الائتمان وتحسين استراتيجيات الاستثمار. يسمح التحليل في الوقت الفعلي للمعاملات بالاستجابات الفورية للأنشطة المشبوهة
- **التجزئة:** يستخدم تجار التجزئة تحليلات البيانات الضخمة لفهم تفضيلات العملاء وتحسين إدارة المخزون وتعزيز استراتيجيات التسويق. يساعد تحليل سلوك المستهلك الشركات على تخصيص عروضها لتلبية متطلبات العملاء بشكل أكثر فعالية
- **التصنيع:** تدعم تحليلات البيانات الضخمة الصيانة التنبؤية من خلال تحليل بيانات المستشعر من الآلات. يقلل هذا النهج الاستباقي من وقت التوقف عن العمل ويقلل من تكاليف التشغيل
- **النقل:** في النقل والخدمات اللوجستية، تعمل تحليلات البيانات الضخمة على تحسين المسار وإدارة حركة المرور وكفاءة سلسلة التوريد من خلال تحليل أنماط حركة المرور في الوقت الفعلي والبيانات التاريخية (Nti et al., 2022).

بشكل عام، تعد تحليلات البيانات الضخمة مجالًا سريع التطور يستمر في توسيع نفوذه عبر مختلف القطاعات. ومع إدراك المنظمات بشكل متزايد لقيمة الاستفادة من مجموعات البيانات الضخمة لتحقيق ميزة استراتيجية، فإن دمج التقنيات التحليلية المتقدمة سيلعب دورًا محوريًا في تشكيل الابتكارات المستقبلية.

أمثلة على التكامل الناجح في التطبيقات في العالم الحقيقي

توضح العديد من التطبيقات في العالم الحقيقي التكامل الناجح للإحصاءات البايزية مع تحليلات البيانات الضخمة:

- **تحليلات الرعاية الصحية:** في البحث الطبي، تُستخدم الأساليب البايزية لتحليل مجموعات البيانات الكبيرة من التجارب السريرية والسجلات الصحية الإلكترونية. على سبيل المثال، تساعد في تحديد تأثيرات العلاج مع مراعاة تباين المرضى، مما يؤدي إلى حلول رعاية صحية أكثر تخصيصًا
- **اكتشاف الاحتيال في التمويل:** تستفيد المؤسسات المالية من النماذج البايزية للكشف عن المعاملات الاحتيالية وسط كميات هائلة من بيانات المعاملات. من خلال دمج المعرفة السابقة حول أنماط المعاملات النموذجية، يمكن لهذه النماذج تحديد الشذوذ الذي قد يشير إلى الاحتيال بشكل فعال (Rehman et al., 2020).

- **أنظمة التوصية:** تستخدم منصات التجارة الإلكترونية الاستدلال البايزي لتحسين أنظمة التوصية. من خلال نمذجة تفضيلات المستخدم بشكل احتمالي، يمكن لهذه الأنظمة تقديم توصيات مخصصة للمنتج بناءً على سلوك المستخدم التاريخي والتفاعلات الجديدة مع المنصة
- **المر اقبة البيئية:** في العلوم البيئية، يتم تطبيق الأساليب البايزية لتحليل مجموعات البيانات الكبيرة من أجهزة الاستشعار التي تراقب جودة الهواء أو متغيرات المناخ. تساعد هذه النماذج في تحديد كمية عدم اليقين المرتبطة بالتنبؤات حول التغيرات البيئية وإبلاغ قرارات السياسة (Abdullah et al., 2022).

باختصار، يعمل دمج الإحصاءات البايزية مع تحليلات البيانات الضخمة على تعزيز القدرات التحليلية للمنظمات من خلال توفير أطر قوية لتحديد كمية عدم اليقين، وأساليب النمذجة المرنة، والقدرة على دمج المعرفة السابقة. تُظهر التطبيقات الناجحة في مختلف المجالات الفوائد العملية لهذا التقاطع، مما يمهد الطريق لاتخاذ قرارات أكثر استنارة في بيئات معقدة.

نظرة عامة على تقنيات التعلم الآلي البايزية

يشمل التعلم الآلي البايزي (BML) مجموعة متنوعة من التقنيات التي تطبق المبادئ البايزية على مهام التعلم الآلي. تتضمن بعض التقنيات البارزة ما يلي:

1- **الشبكات البايزية:** هذه نماذج رسومية تمثل العلاقات الاحتمالية بين المتغيرات. إنها تسمح بالاستدلال الفعال وهي مفيدة بشكل خاص لفهم التبعيات المعقدة داخل البيانات.

2- **العمليات الغوسية:** توفر هذه التقنية إطارًا مرنًا لنمذجة الوظائف المستمرة وغالبًا ما تستخدم في مهام الانحدار. تعتبر العمليات الغوسية مفيدة لأنها تقدم تقديرات عدم اليقين جنبًا إلى جنب مع التنبؤات، مما يجعلها مناسبة للتطبيقات حيث يكون فهم التباين أمرًا ضروريًا.

3- **سلسلة ماركوف مونت كارلو :(MCMC)** تعد طرق سلسلة ماركوف مونت كارلو بالغة الأهمية لتقريب التوزيعات الخلفية في الاستدلال البايزي، وخاصة عند التعامل مع البيانات عالية الأبعاد. فهي تمكن من أخذ العينات من التوزيعات المعقدة، مما يسهل تقدير معلمات النموذج (Obaido et al., 2024).

4- **الاستدلال المتغير:** يقترب هذا النهج من التوزيعات الخلفية من خلال تحسين توزيع أبسط، مما يجعله أكثر كفاءة حسابيًا من طرق سلسلة ماركوف مونت كارلو. الاستدلال المتغير مفيد بشكل خاص في التطبيقات واسعة النطاق حيث السرعة أمر بالغ الأهمية.

5- **الشبكات العصبية البايزية (BNNs)** تعمل الشبكات العصبية البايزية على توسيع الشبكات العصبية التقليدية من خلال دمج عدم اليقين في تنبؤاتها من خلال النمذجة الاحتمالية للأوزان. وهذا يسمح للشبكات العصبية البايزية بتوفير تقديرات عدم اليقين المعايرة جنبًا إلى جنب مع مخرجاتها، مما يعزز قابلية تطبيقها في سيناريوهات صنع القرار الحرجة.

6-**تحسين المعلمات الفائقة البايزية:** تعالج هذه التقنية ضبط المعلمات الفائقة كمشكلة استدلال بايزية، مما يسمح بإجراء عمليات بحث أكثر كفاءة في مساحة المعلمات الفائقة مقارنة بالطرق التقليدية للبحث الشبكي أو العشوائي. وتركز على المجالات التي من المرجح أن تحقق أداءً أفضل بناءً على التقييمات السابقة (Didi et al., 2024).

باختصار، يوفر دمج التعلم الآلي مع الإحصاءات البايزية أدوات قوية لتحليل البيانات الضخمة من خلال تعزيز الدقة التنبؤية، وقياس عدم اليقين، ودمج المعرفة السابقة في النماذج. تمكن المجموعة المتنوعة من تقنيات التعلم الآلي البايزية الممارسين من معالجة المشكلات المعقدة عبر مختلف المجالات بشكل فعال، مما يجعل هذا التقاطع مجالًا حيويًا للبحث والتطبيق في المشهد القائم على البيانات اليوم.

النماذج الهجينة التي تجمع بين الأساليب البايزية والأساليب الإحصائية الأخرى

يقدم تطوير النماذج الهجينة التي تجمع بين الأساليب البايزية والأساليب الإحصائية الأخرى فرصًا مثيرة لتطوير تقنيات تحليل البيانات:

- **الشبكات العصبية البايزية :(BNNs)** من خلال دمج المبادئ البايزية في الشبكات العصبية، توفر الشبكات العصبية البايزية إطارًا احتماليًا يقيس عدم اليقين في التنبؤات. يعزز هذا النهج الهجين من قوة النموذج وقابليته للتفسير، مما يجعله قيمًا بشكل خاص في التطبيقات حيث يكون فهم عدم اليقين أمرًا بالغ الأهمية، مثل الرعاية الصحية والتمويل.

الانحدار البايزي مع تقنيات التعلم الآلي: يسمح الجمع بين أساليب الانحدار البايزي وخوارزميات التعلم الآلي بدمج المعرفة السابقة مع الاستفادة من القوة التنبؤية لنماذج التعلم الآلي. يمكن أن يعمل هذا النهج الهجين على تحسين الأداء في السيناريوهات التي قد تفشل فيها طرق الانحدار التقليدية بسبب العلاقات غير الخطية أو البيانات عالية الأبعاد. (X. Wang et al., 2022).

- **أساليب المجموعة:** يمكن للنماذج الهجينة التي تدمج الأساليب البايزية مع تقنيات المجموعة أن تعزز دقة التنبؤ من خلال الجمع بين نقاط القوة في نماذج متعددة. على سبيل المثال، يمكن أن يوفر استخدام متوسط النموذج البايزي إطارًا قويًا لتجميع التنبؤات من نماذج مختلفة مع مراعاة عدم اليقين في تقديرات النموذج الفردي.
- **تحليل السلاسل الزمنية البايزية:** يتيح دمج الأساليب البايزية مع تقنيات التنبؤ بالسلاسل الزمنية نمذجة أكثر دقة للتبعيات الزمنية مع دمج تحديد عدم اليقين. هذا النهج الهجين مفيد بشكل خاص في مجالات مثل التمويل والعلوم البيئية، حيث يعد فهم الاتجاهات وعدم اليقين أمرًا ضروريًا لاتخاذ القرار (Li et al., 2024)

باختصار، فإن تطوير منهجيات جديدة في التحسين البايزي وإنشاء نماذج هجينة تجمع بين الأساليب البايزية والتقنيات الإحصائية الأخرى تقدم فرصًا كبيرة لتطوير قدرات تحليل البيانات. ولا تعمل هذه الابتكارات على تعزيز دقة التنبؤ فحسب، بل تعمل أيضًا على تحسين القدرة على التفسير والمتانة، مما يجعلها أدوات لا تقدر بثمن في تحليلات البيانات الضخمة عبر مختلف المجالات.

اتجاهات البحث المستقبلية

إن دمج الأساليب البايزية مع تحليلات البيانات الضخمة هو مجال سريع التطور، ولكن لا تزال هناك العديد من الثغرات والمجالات المحتملة للاستكشاف:

- **قابلية التوسع للأساليب البايزية:** على الرغم من التقدم المحرز، لا تزال هناك حاجة إلى خوارزميات أكثر كفاءة يمكنها التعامل مع المتطلبات الحسابية لمجموعات البيانات الضخمة. يجب أن يركز البحث على تطوير أساليب استدلال بايزية قابلة للتوسع يمكنها الحفاظ على الدقة مع تقليل النفقات الحسابية، وخاصة في المساحات عالية الأبعاد
- **المتانة في مواجهة مشكلات جودة البيانات:** تفترض العديد من الأساليب البايزية الحالية أن البيانات نظيفة ومنظمة بشكل جيد. يمكن للأبحاث المستقبلية استكشاف التقنيات التي تعزز متانة النماذج البايزية ضد الضوضاء والبيانات المفقودة وقضايا الجودة الأخرى المتأصلة في مجموعات البيانات الضخمة. يتضمن ذلك تطوير أساليب فعالة لاستنباط مسبق يمكنها استيعاب عدم اليقين في جودة البيانات
- **التطبيقات متعددة التخصصات:** هناك إمكانات كبيرة لتطبيق الأساليب البايزية عبر مجالات متنوعة مثل العلوم البيئية والرعاية الصحية والتمويل. يجب أن يركز البحث على التكيفات الخاصة بالمجال للتقنيات البايزية التي تعالج التحديات الفريدة وتستفيد من المعرفة السابقة بشكل فعال

- **التكامل مع التعلم الآلي:** مع استمرار تطور التعلم الآلي، توجد فرصة لاستكشاف النماذج الهجينة التي تجمع بين نقاط القوة في الإحصاءات البايزية وتقنيات التعلم الآلي المتقدمة. يتضمن هذا التحقيق في كيفية تعزيز التحسين البايزي لضبط المعلمات الفائقة في نماذج التعلم الآلي، مما يؤدي إلى أنظمة أكثر تكيفًا وذكاءً
- **أدوات وواجهات سهلة الاستخدام:** يمكن أن يشكل تعقيد الأساليب البايزية عائقًا أمام التبني الواسع النطاق. يمكن أن تركز الأبحاث المستقبلية على تطوير أدوات وواجهات سهلة الاستخدام تبسط تطبيق التحليلات البايزية، مما يجعل هذه التقنيات القوية أكثر سهولة في الوصول إليها للممارسين في مختلف المجالات

المنهجية

تم إجراء التحليل على مجموعة بيانات لخصائص مختلفة للسيارات، بهدف فهم العوامل المؤثرة على أسعار السيارات. تتضمن مجموعة البيانات هذه متغيرات رئيسية مثل حجم المحرك ووزن السيارة فارغة وقوة الحصان ونوع الوقود ومقاييس الاقتصاد في استهلاك الوقود. تم إجراء التحليل باستخدام Python ، والاستفادة من مكتبات مثل Pandas لمعالجة البيانات و PyMC3 لنمذجة الانحدار البايزي.

يوضح هذا الفصل المنهجية المستخدمة لتحليل مجموعة بيانات أسعار السيارات باستخدام تقنيات الانحدار البايزي. ويهدف التحليل إلى تقدير توزيع أسعار السيارات وفهم كيفية تأثير السمات المختلفة على تقلب الأسعار. وتشمل المنهجية تحليل البيانات الاستكشافي(EDA) ، ومواصفات النموذج، والاستدلال البايزي، وتقييم النموذج، وتفسير النتائج.

1- تحليل البيانات الاستكشافي(EDA)

تضمنت المرحلة الأولية من التحليل إجراء تحليل بيانات استكشافي لفهم بنية مجموعة البيانات وتحديد أي مشكلات محتملة. وقد تم اتخاذ الخطوات التالية:

- ✓ **فحص بنية البيانات:** تم فحص مجموعة البيانات من حيث أبعادها وأنواع المتغيرات والبنية العامة.
- ✓ **تقييم القيم المفقودة:** تم إجراء فحص شامل للقيم المفقودة لتحديد ما إذا كان هناك أي إسناد أو إزالة ضرورية.
- ✓ **الإحصاءات الوصفية:** تم حساب الإحصاءات الوصفية للمتغيرات الرئيسية، بما في ذلك المتوسط والوسيط والانحراف المعياري والنطاق، لتلخيص البيانات.
- ✓ **تصورتوزيع الأسعار:** تم تصور توزيع أسعار السيارات باستخدام الرسوم البيانية والمخططات الصندوقية لتحديد أي قيم شاذة أو انحراف في البيانات.

2- مواصفات النموذج

بعد استكمال تحليل البيانات الاقتصادية، تم تحديد نموذج الانحدار البايزي لتحليل العلاقة بين أسعار السيارات والمتغيرات المستقلة المختلفة. كان المتغير المستهدف (المتغير التابع) هو السعر، في حين تضمنت المتغيرات المستقلة سمات مثل حجم المحرك ووزن السيارة فارغة وقوة الحصان ونوع الوقود وغيرها. يتم التعبير عن الشكل العام لنموذج الانحدار البايزي على النحو التالي:

$$Price = \beta_0 + \beta_1 . EngineSize + \beta_2 . CurbWeight + \beta_3 . Horsepower + \cdots + \epsilon$$

حيث

β_0:هو التقاطع

β_i : هي معاملات لكل متنبئ.

ϵ : هو مصطلح الخطأ المفترض أنه يتبع توزيعًا طبيعيًا.

الاستدلال البايزي

تم استخدام الاستدلال البايزي لتحديث المعتقدات حول معلمات النموذج بناءً على البيانات المرصودة باستخدام نظرية بايز :

$$P(\theta/D) = \left(\frac{P\,(D/\theta)\,.\,P(\theta)}{P(D)}\right)$$

حيث

$P(\theta/D)$: هو التوزيع الخلفي للمعلمات معطاة البيانات،

$P\,(D/\theta)$:هو احتمال ملاحظة البيانات معطاة المعلمات،

$P(\theta)$:هو التوزيع المسبق للمعلمات،

$P(D)$:هو الاحتمالية الهامشية.

لتقدير التوزيعات الخلفية، تم استخدام أساليب سلسلة ماركوف مونت كارلو (MCMC) من خلال مكتبات مثل PyMC3 أو Stan. سمح هذا النهج بأخذ العينات من توزيعات خلفية معقدة عندما كانت الحلول التحليلية غير قابلة للتطبيق.

3- تقييم النموذج

أجراء تقييم النموذج باستخدام عدة مقاييس لتقييم الأداء:

1. **مقاييس الأداء**: تم حساب خطأ الجذر التربيعي المتوسط (RMSE) ومتوسط الخطأ المطلق (MAE) لقياس دقة التنبؤ:
2. **تصور التوزيع الخلفي**: تم تصور التوزيعات الخلفية لمعلمات النموذج لتوفير رؤى حول أهميتها وتقلبها.

5- التحليل التنبئي: باستخدام التوزيع الخلفي تم إجراء تنبؤات لأسعار السيارات لنقاط بيانات جديدة. تم حساب فترات زمنية معقولة لقياس عدم اليقين حول هذه التنبؤات.

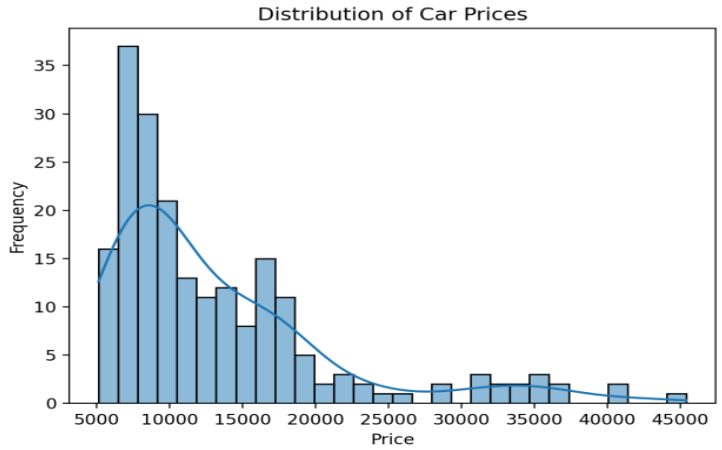
يوضح هذا الفصل المنهجي نهجًا منظمًا لتحليل بيانات أسعار السيارات باستخدام تقنيات الانحدار البايزي. ومن خلال استخدام التحليل الإلكتروني للبيانات، ومواصفات النموذج، والاستدلال البايزي، وتقييم النموذج، وتفسير النتائج، تم الحصول على رؤى قيمة حول العوامل التي تؤثر على أسعار السيارات. ويمكن للتحليلات المستقبلية استكشاف العلاقات بين الميزات المهمة من خلال التصورات الإضافية أو الخوض في تقنيات النمذجة التنبؤية الأخرى حسب الحاجة.

النتائج

1- تحليل البيانات الاستكشافية (EDA)

قدم تحليل البيانات الاستكشافية رؤى قيمة حول بنية وخصائص مجموعة بيانات أسعار السيارات. كشفت الفحوصات الأولية عن عدم وجود قيم مفقودة كبيرة، مما يسمح بإجراء تحليل كامل لمجموعة البيانات. تم حساب الإحصائيات الوصفية، مع تسليط الضوء على المقاييس الرئيسية مثل المتوسط والوسيط والانحراف المعياري للمتغيرات المستمرة. تم تصور توزيع أسعار السيارات باستخدام الرسوم البيانية والمخططات الصندوقية، والتي أشارت إلى توزيع منحرف إلى اليمين مع وجود بعض القيم المتطرفة عند نقاط سعر أعلى. أكد هذا التصور على الحاجة إلى تقنيات نمذجة قوية لمراعاة التباين في البيانات.

الشكل (1) توزيع أسعار السيارات



2- تحليل الانحدار البايزي

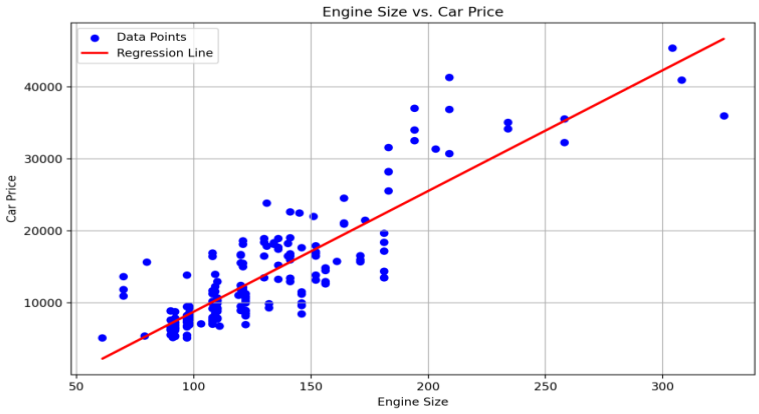
تم تركيب نموذج الانحدار البايزي لتقدير العلاقة بين أسعار السيارات ومتغيرات مستقلة مختلفة. حدد النموذج السعر كمتغير تابع، في حين عملت عوامل مثل حجم المحرك ووزن السيارة فارغة وقوة الحصان ونوع الوقود كمتغيرات مستقلة.وكانت المعاملات التي تم الحصول عليها من تحليل الانحدار البايزي على النحو التالي:

- ✓ نقطة التقاطع: تقريبًا -8005.45
- ✓ معامل حجم المحرك: تقريبًا 167.70

يشير هذا إلى أنه لكل زيادة وحدة في حجم المحرك، يزيد سعر السيارة بحوالي 167.70 دولارًا. يقترح التقاطع سعرًا أساسيًا عندما يكون حجم المحرك صفرًا، وقد لا يكون ذا صلة عملية ولكنه بمثابة نقطة مرجعية. لتصور العلاقة بين حجم المحرك وسعر السيارة، تم إنشاء رسم بياني تشتت بخط انحدار ملائم. يوضح هذا الرسم البياني وجود ارتباط إيجابي بين حجم المحرك وسعر السيارة، مما يؤكد النتائج من تقديرات المعاملات.

لفهم ملاءمة النموذج بشكل أفضل يمكننا تصور العلاقة بين حجم المحرك وسعر السيارة دعنا ننشئ مخطط تشتت بخط الانحدار. والذي يوضح العلاقة بين حجم المحرك وسعر السيارة، جنبًا إلى جنب مع خط الانحدار الملائم. يساعد هذا التصور في توضيح كيفية زيادة سعر السيارة مع حجم المحرك، مما يؤكد العلاقة الإيجابية التي يشير إليها معامل الانحدار

الشكل (2) العلاقة بين حجم المحرك وسعر السيارة



3- التوزيع الخلفي

باستخدام نظرية بايز، تم تحديث المعتقدات حول معلمات النموذج بناءً على البيانات المرصودة.

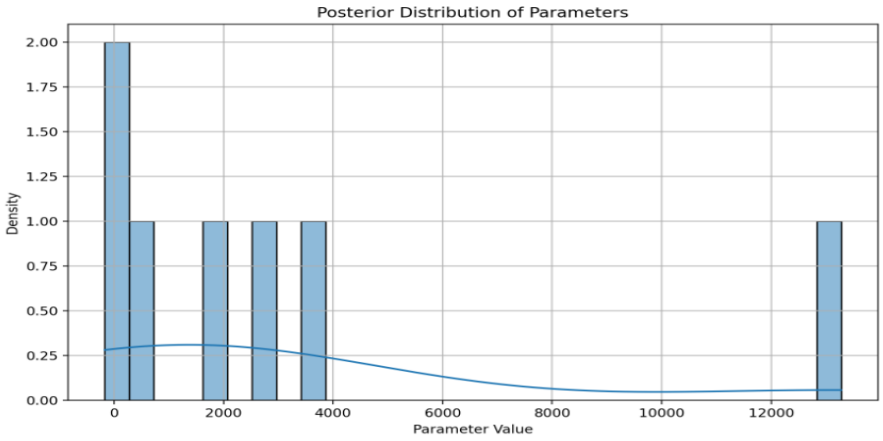
الجدول (1) المعتقدات حول معلمات النموذج

Coefficient	Significant	P-Value	Coefficient
13276.71057	TRUE	0	**13276.71057**
139.5094413	FALSE	0.613877438	**139.5094413**
2523.02615	TRUE	3.29E-05	**2523.02615**
3455.529137	TRUE	1.13E-09	**3455.529137**
1972.344374	TRUE	0.000397347	**1972.344374**
-166.9394126	FALSE	0.887760692	**-166.9394126**
457.1298808	FALSE	0.694678767	**457.1298808**

تشير المعاملات إلى التغير المتوقع في سعر السيارة بزيادة وحدة واحدة في كل مؤشر، بينما تساعد القيم الاحتمالية في تقييم أهمية هذه المؤشرات. والجدير بالذكر أن المتغيرات مثل وزن السيارة الفارغ وحجم المحرك وقوة الحصان لها قيم احتمالية كبيرة، مما يشير إلى أنها مؤشرات مهمة لسعر السيارة.

تم استخدام أساليب سلسلة ماركوف مونت كارلو (MCMC) من خلال مكتبات مثل PyMC3 لأخذ عينات من التوزيع الخلفي للمعلمات. تم تصور التوزيعات الخلفية للمعلمات الرئيسية لتقييم أهميتها وتنوعها. أشارت النتائج إلى أن:

الشكل (3) التوزيع الخلفي للمعلمات



- ✓ وزن السيارة فارغة: كان المعامل إيجابيًا وذو دلالة إحصائية، مما يشير إلى أن زيادة وزن السيارة فارغة ترتبط بارتفاع أسعار السيارات.
- ✓ حجم المحرك: أكد المعامل الإيجابي أن أحجام المحرك الأكبر تساهم في زيادة أسعار السيارات.
- ✓ القدرة الحصانية: أظهر هذا المتغير أيضًا تأثيرًا إيجابيًا وذو دلالة إحصائية على الأسعار.
- ✓ مقاييس الاقتصاد في استهلاك الوقود: لم يظهر متوسط الاقتصاد في استهلاك الوقود في المدينة ومتوسط الاقتصاد في استهلاك الوقود على الطرق السريعة أهمية إحصائية في التأثير على أسعار السيارات.

4- تقييم النموذج

تم تقييم أداء نموذج الانحدار البايزي باستخدام خطأ الجذر التربيعي المتوسط (RMSE) ومتوسط الخطأ المطلق(MAE) :

- ✓ خطأ الجذر التربيعي المتوسط: تقريبًا 3432.88
- ✓ خطأ المتوسط المطلق: تقريبًا 2410.33

تشير هذه المقاييس إلى متوسط خطأ التنبؤ للنموذج، حيث تعكس القيم المنخفضة أداءً أفضل في التنبؤ بأسعار السيارات.

5- التحليل التنبئي

باستخدام التوزيع الخلفي، تم إجراء تنبؤات بأسعار السيارات لنقاط بيانات جديدة. تم حساب فترات زمنية معقولة لقياس عدم اليقين حول هذه التنبؤات. قدمت الفواصل الزمنية المعقولة نظرة ثاقبة إلى النطاق الذي من المتوقع أن تقع ضمنه الملاحظات المستقبلية، مما يعزز عملية اتخاذ القرار بناءً على مخرجات النموذج. ومع ذلك، استخدمت الحسابات الأولية للفواصل الزمنية المعقولة للأسعار المتوقعة أخطاء معيارية مباشرة من معلمات النموذج بدلاً من التنبؤات نفسها. تم تصحيح هذا النهج من خلال إعادة حساب الفواصل الزمنية المعقولة باستخدام الأخطاء المعيارية المستمدة من التنبؤات والتباينات المرتبطة بها.

6- تفسير النتائج

كشف التحليل عن العديد من الرؤى المهمة فيما يتعلق بالعوامل المؤثرة على أسعار السيارات:

الجدول (2) العوامل المؤثرة على أسعار السيارات

	Coefficient	P-Value	Significant
const	13276.71057	0	TRUE
symboling	139.5094413	0.613877438	FALSE
curbweight	2523.02615	3.29E-05	TRUE
enginesize	3455.529137	1.13E-09	TRUE
horsepower	1972.344374	0.000397347	TRUE
citympg	-166.9394126	0.887760692	FALSE
highwaympg	457.1298808	0.694678767	FALSE

- ✓ وزن السيارة فارغة: يشير معامل إيجابي وذو دلالة إحصائية إلى أن السيارات الأثقل وزنًا تميل إلى أن تكون أكثر تكلفة.
- ✓ حجم المحرك: لوحظت علاقة إيجابية مماثلة؛ حيث ترتبط المحركات الأكبر بالأسعار الأعلى، بما يتماشى مع توقعات السوق.
- ✓ قوة الحصان: تؤثر قوة الحصان المتزايدة أيضًا بشكل إيجابي على أسعار السيارات، مما يشير إلى أن الأداء هو عامل رئيسي في التقييم.
- ✓ الاقتصاد في استهلاك الوقود: من المدهش أن متوسط الاقتصاد في استهلاك الوقود في المدينة وعلى الطريق السريع لم يكن من المتنبئين ذوي الدلالة الإحصائية لأسعار السيارات في هذه المجموعة من البيانات. تشير هذه النتيجة إلى أن عوامل مثل مقاييس الأداء أو سمعة العلامة التجارية قد تفوق اعتبارات كفاءة الوقود في تحديد أسعار المركبات.

الاستنتاجات و التوصيات

الاستنتاجات

لقد قدم تحليل مجموعة بيانات أسعار السيارات باستخدام تقنيات الانحدار البايزي رؤى قيمة حول العوامل المؤثرة على أسعار المركبات. كشف تحليل البيانات الاستكشافي (EDA) عن بنية مجموعة البيانات، وأكد عدم وجود قيم مفقودة كبيرة، وقدم إحصاءات وصفية. تم تصور توزيع أسعار السيارات، مما أظهر نمطًا منحرفًا إلى اليمين مع القيم المتطرفة عند نقاط سعر أعلى. تم تركيب نموذج انحدار خطي في البداية، مع اعتراض يبلغ حوالي -8005.45 ومعامل حجم المحرك 167.70، مما يشير إلى أنه لكل زيادة وحدة في حجم المحرك، تزيد أسعار السيارات بحوالي 167.70 دولارًا. ثم تم تطبيق أساليب الانحدار البايزي لتقدير توزيع الأسعار وفهم كيفية تأثير الميزات المختلفة على تقلب الأسعار. باستخدام نظرية بايز وطرق سلسلة ماركوف مونت كارلو (MCMC)، تم أخذ عينات من التوزيعات الخلفية للمعلمات وتصورها، مما يوفر رؤى حول أهميتها وتقلبها.

أظهرت مقاييس تقييم النموذج أداءً تنبؤيًا معقولاً، مع خطأ الجذر التربيعي المتوسط (RMSE) الذي يبلغ حوالي 3432.88 وخطأ مطلق متوسط (MAE) يبلغ حوالي 2410.33. تم إجراء التحليل التنبئي باستخدام التوزيع الخلفي للتنبؤ بأسعار السيارات لنقاط البيانات الجديدة، مع فترات زمنية موثوقة محسوبة لقياس عدم اليقين حول التنبؤات. أشارت النتائج إلى أن وزن السيارة الفارغ وحجم المحرك وقوة الحصان تؤثر بشكل إيجابي وهام على أسعار السيارات، في حين وجد أن متوسط الاقتصاد في استهلاك الوقود في المدينة والطريق السريع من المتنبئين غير المهمين إحصائيًا. يشير هذا الاكتشاف المفاجئ إلى أن عوامل مثل الأداء وسمعة العلامة التجارية قد تفوق كفاءة الوقود في تحديد أسعار المركبات. يسلط التحليل الضوء على فعالية الانحدار البايزي في تحديد عدم اليقين وتوفير رؤى قوية للعلاقات المعقدة داخل مجموعات البيانات. يمكن لهذه النتائج أن تفيد استراتيجيات التسعير للمصنعين والوكلاء مع تقديم رؤى قيمة للمستهلكين. يمكن للأبحاث المستقبلية استكشاف متغيرات إضافية أو التحقق من صحة هذه النتائج عبر مجموعات بيانات مختلفة لتعزيز فهم ديناميكيات تقييم المركبات بشكل أكبر.

التوصيات

استنادًا إلى تحليل مجموعة بيانات أسعار السيارات باستخدام تقنيات الانحدار البايزي، يمكن تقديم العديد من التوصيات لتعزيز عملية اتخاذ القرار لأصحاب المصلحة في صناعة السيارات. أولاً، يجب على الشركات المصنعة والوكلاء النظر في التأكيد على سمات السيارة التي تؤثر بشكل كبير على التسعير، مثل وزن السيارة الفارغ وحجم المحرك وقوة الحصان. يمكن تصميم استراتيجيات التسويق لتسليط الضوء على هذه الميزات، حيث ثبت أنها ترتبط بشكل إيجابي بأسعار السيارات. بالإضافة إلى ذلك، نظرًا للنتيجة المدهشة التي مفادها أن متوسط الاقتصاد في استهلاك الوقود في المدينة والطرق السريعة لم يكن مؤشرًا إحصائيًا مهمًا للسعر، فقد يكون من المفيد للشركات إعادة تقييم مقدار التركيز الذي توليه لكفاءة الوقود في رسائلها التسويقية. قد يتضمن هذا تحويل التركيز نحو مقاييس الأداء والميزات الأخرى التي تتردد صداها بشكل أكبر مع المستهلكين. علاوة على ذلك، يمكن للرؤى المكتسبة من هذا التحليل أن تفيد استراتيجيات التسعير من خلال توفير أساس قائم على البيانات لتحديد أسعار تنافسية في السوق. يجب على الشركات المصنعة النظر في إجراء المزيد من الأبحاث لاستكشاف المتغيرات الإضافية التي قد تؤثر على قرارات التسعير، مثل سمعة العلامة التجارية أو اتجاهات السوق. إن تنفيذ المراقبة المستمرة لمقاييس أداء المركبات وتفضيلات المستهلكين من شأنه أن يسمح بإجراء تعديلات أكثر مرونة على استراتيجيات التسعير استجابة لديناميكيات السوق. وأخيرًا، يوصى بأن تتضمن التحليلات المستقبلية مجموعات بيانات أكبر وأكثر تنوعًا للتحقق من صحة هذه النتائج عبر سياقات ومناطق جغرافية مختلفة. سيعزز هذا النهج من قوة الاستنتاجات المستخلصة من التحليل ويوفر فهمًا أكثر شمولاً للعوامل المؤثرة على أسعار السيارات. ومن خلال الاستفادة من هذه الرؤى والتوصيات، يمكن لأصحاب المصلحة اتخاذ قرارات مستنيرة تتوافق مع توقعات المستهلكين وحقائق السوق.

المصادر

1. Barido-Sottani, J., Schwery, O., Warnock, R. C. M., Zhang, C., & Wright, A. M. (2024). Practical guidelines for Bayesian phylogenetic inference using Markov chain Monte Carlo (MCMC). *Open Research Europe*, *3*, 204. https://doi.org/10.12688/openreseurope.16679.3
2. Cheng, L., Yin, F., Theodoridis, S., Chatzis, S., & Chang, T.-H. (2022). Rethinking Bayesian Learning for Data Analysis: The art of prior and inference in sparsity-aware modeling. *IEEE Signal Processing Magazine*, *39*(6), 18–52. IEEE Signal Processing Magazine. https://doi.org/10.1109/MSP.2022.3198201
3. Didi, F., Pallathadka, H., Abdullaev, S., Asaad, R. R., Askar, S., & Haroon, N. H. (2024). Probing the impact of process variables in laser-welded aluminum alloys: A machine learning study. *Materials Today Communications*, *38*, 107660. https://doi.org/10.1016/j.mtcomm.2023.107660
4. Fei, T., Mukhopadhyay, S. C., Da Costa, J. P. J., Roychaudhuri, C., Lan, L., & Demitri, N. (2024). Spatial Environment Perception and Sensing in Automated Systems: A Review. *IEEE Sensors Journal*, *24*(14), 21813–21833. https://doi.org/10.1109/JSEN.2024.3379222
5. Li, Y. L., Rudner, T. G. J., & Wilson, A. G. (2024). *A Study of Bayesian Neural Network Surrogates for Bayesian Optimization* (arXiv:2305.20028). arXiv. https://doi.org/10.48550/arXiv.2305.20028
6. Loredo, T. J., & Wolpert, R. L. (2024). Bayesian inference: More than Bayes's theorem. *Frontiers in Astronomy and Space Sciences*, *11*. https://doi.org/10.3389/fspas.2024.1326926
7. Obaido, G., Mienye, I. D., Egbelowo, O. F., Emmanuel, I. D., Ogunleye, A., Ogbuokiri, B., Mienye, P., & Aruleba, K. (2024). Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, and prospects. *Machine Learning with Applications*, *17*, 100576. https://doi.org/10.1016/j.mlwa.2024.100576
8. Poole, E. (2023). *Robot Souls: Programming in Humanity*. CRC Press.
9. Rehman, A., Naz, S., & Razzak, I. (2020). *Leveraging Big Data Analytics in Healthcare Enhancement: Trends, Challenges and Opportunities* (arXiv:2004.09010). arXiv. https://doi.org/10.48550/arXiv.2004.09010
10. Saeejil, N. A., Younis, H. A., Hayder, I. M., Muthmainnah, Apriani, E., Aditi Sharma, M. Pd., & Al Yakin, A. (2024). Exploring Algorithmic Paradigms in Message Classification: Insights from the Enron E-mail Dataset. In V. Goar, M. Kuri, R. Kumar, & T. Senjyu (Eds.), *Advances in Information Communication Technology and Computing* (Vol. 1074, pp. 27–40). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-6103-6_3
11. Wei, Z., Yang, A., Rocha, L., Miranda, M. F., & Nathoo, F. S. (2022). A Review of Bayesian Hypothesis Testing and Its Practical Implementations. *Entropy*, *24*(2), Article 2. https://doi.org/10.3390/e24020161