

خلال العقد الأخير أصبح انخفاض الرؤية يمثل تحديا هاما للمتنبئين بالطقس، بالنظر إلى تأثيره السلبي على حركة السير وحركة الملاحة الجوية والبحرية. في الواقع أصبحت الخسائر البشرية والمالية التي تعزى إلى انخفاض الرؤية جد وخيمة، وبالتالي من أجل مواجهة هذا العجز في مجال التنبؤ العددي بالطقس تم تقييم تقنيات Ddatamining لتقدير والتنبؤ بالرؤية الأفقية في العديد من الدراسات العلمية. إلا أن نتائج أداء وقدرة النماذج المطورة تختلف من دراسة لأخرى بسبب تنوع الأدوات Datamining المفتوحة المصدر المستعملة من جهة، وتنوع الخوارزميات Datamining المستخدمة من جهة أخرى.

لذلك فالهدف من دراستنا هو دراسة حساسية أداء وقدرة النماذج المطورة لأداة ولخوارزمية Datamining، دراسة حالة Régression في تنبؤ الرؤية انطلاقا من التنبؤات الرقمية ل AROME. لتحقيق هذا الهدف استخدمنا نوعين من الخوارزميات، تلك القائمة على Ensemble methods بما فيها Random Forest، Gradient Boosting Machine، eXtreme Gradient Boosting وغيرها القائمة على التعلم العميق Deep Learning. هذه التقنيات تم تقييمها في مختلف المنصات المفتوحة المصدر Scikit-learn، H2O، WEKA، Tensorflow، Keras. بالإضافة إلى ذلك، تم استخدام في هذا العمل قاعدة بيانات تغطي بيانات لحظية لمدة 3 سنوات كنتيجة للمعالجة المسبقة لمخرجات نموذج التنبؤ العددي AROME وبيانات مرصودة بالملاحظة. تم تقسيم هذه البيانات إلى 70% للتعلم و 30% للاختبار مع الحرص على تمثيلية كل شهر، كل ساعة، ومختلف مجالات الرؤية لجميع المحطات.

تظهر نتائج تقييم حساسية النماذج التي تم تطويرها للأداة والخوارزمية المستخدمة أن أداء النماذج المستندة على Ensemble methods هو الأفضل مهما كانت الأداة المستخدمة باستثناء Keras حيث تم استخدامه للتعلم العميق فقط. من ناحية أخرى، يتم عرض خوارزمية Random Forest كأفضل مقدر للرؤية بعد ضبط إعدادات hyperparameter لمنصاتي WEKA و Scikit-learn. ومع ذلك، Gradient Boosting Machine تميزت بأنها أفضل مقدر لمنصة H2O. نسب الأخطاء المسجلة متشابهة بين مختلف المنصات، إذ نجد أن متوسط الأخطاء التربيعية هو 1933م و 1942 م و 1945 م على التوالي ل Gradient Boosting في H2O و Random Forest لمنصتي Scikit-learn و WEKA.