

Département informatique

Mémoire de projet de fin d'études

Présenté par : Ayoub Bouihrouchane

Pour l'obtention de :

Master de recherche sciences des données

Thème:

**Prediction de débit au sous bassin de Rheraya
à l'aide des modèles d'intelligence artificielle :
Recherche et déploiement**

Encadré par :

Mr. Soufiane IDBRAIM (FSA)

Mme. Salwa BELAQZIZ (FSA – CRSA)

Mr. Abdelghani BOUDHAR (CRSA)

Mr. Haytam ELYOUSSFI (CRSA)

Soutenu le 4 Juillet 2022

Année universitaire 2021-2022

Dédicaces

A mes très chers parents, vous êtes la source de ma force et ma passion, pour vos prières, votre encouragement et votre soutien. Puisse Dieu, le tout puissant vous protège de tout mal et vous accorde santé et miséricorde.

A mes frères et sœurs, rien ne pourra exprimer mes gratitude, mon respect et mon amour éternel que j'ai pour vous.

A ma grande famille, je ne trouve pas les mots pour vous remercier pour tout ce que vous avez fait pour moi.

A mes amis, mes professeurs, et tous ceux qui ont sacrifié pour moi, trouvez dans ce travail l'expression de mon profond amour, je dédie ce travail à vous tous.

Remerciements

Avant toute chose, je tiens à remercier dans un premier temps, le Dieu pour cette grâce d'être en vie et en bonne santé depuis ma naissance jusqu'à ce jour, pour tout ce que j'ai et ce que je suis.

Je tiens encore à exprimer mes profondes gratitude et mes sincères remerciements à mes professeurs encadrant **monsieur Soufiane IDBRAIM** et **Madame Salwa BELAQZIZ** pour leur supervision, leur encadrement, leurs précieux conseils, leurs suggestions de lecture, leurs corrections et leurs qualités scientifiques pour mener bien ce travail.

J'adresse ainsi mes sincères remerciements à **monsieur Abdelghani BOUDHAR** et **monsieur Haytam ELYOUSSFI** pour le temps qu'ils m'ont réservé, pour le soutien qu'ils m'ont fourni, et pour m'avoir fait partager leurs expériences et leurs compétences.

Je tiens également à exprimer mes sincères remerciements à **l'équipe MorSnow** et à l'ensemble des ingénieurs et doctorants du **CRSA** pour l'accueil chaleureux dont j'ai bénéficié, pour leurs explications et informations fournies...

Enfin, je ne peux pas achever ce projet sans exprimer mes gratitude à notre coordinateur du master **monsieur Karim AFDEL** et à tous les enseignants du département informatique de la faculté des sciences d'Agadir, pour leur dévouement et leur assistance tout au long de mes études.

Résumé

Ce Projet intitulé « **Prediction de débit au sous bassin de Rheraya à l'aide des modèles d'intelligence artificielle : Recherche et déploiement** » est réalisé dans le cadre de la préparation du projet fin d'étude au sein du centre CRSA (Center for Remote Sensing Application) à l'université Mohammed VI polytechnique dans le cadre d'un programme de recherche MorSnow.

Ce travail a pour objectif la création et le déploiement des modèles Machine Learning et Deep Learning dans une plateforme web considérée comme un outil d'aide à la décision, qui sera mise à la disposition des gestionnaires pour quantifier les ressources en eau et faire des prédictions des débit d'eau passant par un bassin versant afin de concevoir des plans de gestion des ressources en eau et anticiper les événements hydrologiques (sécheresse, crues...).

Ce projet a focalisé sur le bassin de Tensift qui contient un sous bassin pilote qui s'appelle Rheraya, notre objectif est de créer des modèles Machine learning et deep Learning assez performants pour prédire le débit d'eau dans le sous bassin de Rheraya, dans ce projet nous avons intégré le modèle conceptuel SnowMelt Runoff Model (SRM) pour qu'on puisse comparer les performances des modèles d'intelligence artificielle par rapport aux modèles hydrologiques.

Dans ce travail, nous avons comparé 3 modèles : Support Vector Regressor (SVR), Long Short Term Memory (LSTM), eXtreme Gradient Boosting (XGBoost), avec différents scénarios et différents inputs afin de trouver le modèle le plus performant avec l'utilisation de la méthode de dichotomie et la descente du gradient pour trouver les bonnes combinaisons des hyper paramètres. Les entrées sont des données tabulaires (données météorologiques, données satellitaires...), ces données sont des données journalières collectées entre 2003 et 2016.

Mot clés : débit d'eau, bassin du Tensift, Machine Learning, Deep Learning, modèles hydrologiques, dichotomie, descente du gradient

Abstract

This project entitled « **Streamflow prediction in Rheraya sub basin using artificial intelligence methods: research and deployment** » is carried out as part of the preparation of the end of study project in Center for Remote Sensing Applications in Mohammed VI Polytechnic University as part of a MorSnow research program.

The objective of this project is to create and deploy machine learning and deep learning models in a web platform as a decision-making tool. Which will be used by managers to quantify water resources and predict the streamflow through a catchment area in order to design water resource management plans and anticipate hydrological events (drought, floods, etc...)

This project focused on the Tensift basin which contains a pilot sub-basin called Rheraya. Our objective is to create machine learning and deep learning models quite efficient to predict the streamflow in Rheraya Sub bassin, in this project I integrated the conceptual model « SnowMelt Runoff Model » (SRM) in order to compare the performance of artificial intelligence models with hydrological models.

In this project, I compared 3 models: Support Vector Regressor (SVR), Long Short Term Memory (LSTM), eXtreme Gradient Boosting (XGBoost) with different scenarios and different inputs in order to find the most performant model with the use of dichotomy and gradient descent algorithms to find the best hyper parameters combinations. Models inputs were tabular data (meteorological data, satellite data ...) which were collected between 2003 and 2016.

Keyword: Streamflow, Tensift basin, Machine Learning, Deep Learning, hydrological models, dichotomy, gradient descent.

Table des matières

INTRODUCTION GENERALE	1
PARTIE I : CONTEXTE GENERAL, METHODES ET MATERIELS ET ANALYSE DES DONNEES 3	
1 CONTEXTE GENERAL	4
Introduction.....	4
1.1 Problématique	4
1.2 Projet Morsnow.....	5
1.3 Objectif du projet de fin d'étude.....	5
1.4 Plan du projet	6
Conclusion	6
2 METHODES ET MATERIELS.....	7
Introduction.....	7
2.1 Zone d'étude	7
2.2 Les données collectées	8
2.2.1 Les données des stations météorologiques	8
2.2.2 Les données satellitaires	10
2.2.3 Les données de réanalyse.....	11
2.3 Les modèles d'intelligence artificielle	12
2.3.1 Support Vector Regressor	13
2.3.2 Long Short Term Memory	14
2.3.3 EXtreme Gradient Boosting.....	15
2.4 Les modèles hydrologiques.....	16
2.4.1 Snowmelt Runoff Model.....	16
2.5 Réglages des hyper paramètres	17

Conclusion	17
3 ANALYSE DES DONNEES.....	18
Introduction.....	18
3.1 Exploration des données	18
3.1.1 Précipitation	18
3.1.2 Température	19
3.1.3 Humidité	20
3.1.4 Snow Water Equivalent	21
3.1.5 Snowmelt	23
3.1.6 Snow Cover Area.....	24
3.1.7 Débit.....	27
3.2 Feature selection	28
Conclusion	29
PARTIE II : EXPERIENCES ET DEPLOIEMENT.....	30
EXPERIENCES	31
Introduction.....	31
4.1 Prétraitement des données.....	31
4.1.1 Standardisation.....	31
4.1.2 Normalisation Min-Max	32
4.2 Les métriques	33
4.3 Réglage des hyper paramètres	33
4.3.1 Problème	33
4.3.2 Méthode de dichotomie.....	35
4.3.3 Descente du gradient.....	36
4.3.4 Comparaison avec Grid search	38
4.4 Création des modèles	39
4.4.1 Data split	39

4.4.1 Architectures	39
4.4.2 Résultats	42
4.4.3 Discussion	44
4.5 Modèles existants	45
Conclusion	46
5 DEPLOIEMENT	47
Introduction.....	47
5.1 La conception de la plateforme.....	47
5.1.1 Les besoins fonctionnels	47
5.1.2 Architecture de la plateforme.....	47
5.1.3 Diagramme de navigation	48
5.2 L'environnement logiciel.....	49
5.2.1 Base de données	49
5.2.2 Langage.....	50
5.2.3 Framework	51
5.3 Les interfaces	52
5.3.1 Page d'accueil	52
5.3.1 Bloc des données de réanalyse.....	52
5.3.2 Bloc des modèles	53
5.3.3 SRM.....	59
Conclusion	62
CONCLUSION GENERALE ET PERSPECTIVES	63
BIBLIOGRAPHIE.....	64

Liste des figures

Figure 1: Diagramme de GANTT	6
Figure 2: Site du bassin de Tensift et ses sous bassins	8
Figure 3: l'emplacement des stations au sein du sous bassin de Rheraya.....	9
Figure 4: le principe de réanalyse	12
Figure 5: comparaison entre les noyaux de SVR.....	14
Figure 6: architecture LSTM	15
Figure 7: le principe de XGBoost	16
Figure 8: méthode de grid search.....	17
Figure 9: Série des précipitations au sous bassin de Rheraya entre 2003 et 2016.....	18
Figure 10: Température au sous bassin de Rheraya entre 2003 et 2016.....	19
Figure 11: La moyenne annuelle de la température au sous bassin de Rheraya entre 2003 et 2016.....	20
Figure 12: Humidité au sous bassin de Rheraya entre 2003 et 2016	21
Figure 13: Série des valeurs de SWE entre 2003 et 2016.....	22
Figure 14: Comparaison de la temperature mensuelle moyenne avec SWE	22
Figure 15: Snowmelt au sous de Rheraya entre 2003 et 2016.....	23
Figure 16: La relation entre Snowmelt et SWE	24
Figure 17: Snow Cover Area au sous bassin de Rheraya entre 2003 et 2016	25
Figure 18: Matrice de corrélation entre les moyennes mensuelles de SCA, SWE et Snowmelt	26
Figure 19: Les moyennes mensuelles de SCA, SWE et Snowmelt entre 2003 et 2016	26
Figure 20: Le débit journalier et mensuel au sous bassin de Rheraya entre 2003 et 2016	27
Figure 21: La matrice de corrélation entre l'ensemble des variables	28
Figure 22: Auto corrélation en fonction du temps de latence.....	29
Figure 23: l'effet de la standardisation	31
Figure 24: La boîte à moustache des variables humidité et température avant la normalisation	32
Figure 25: l'effet de la normalisation sur les variables humidité et température.....	32
Figure 26: L'évolution de R^2 en fonction de nombre d'estimateurs du modèle XGBoost.....	34
Figure 27: l'évolution de R^2 en fonction de la valeur de C du modèle SVR.....	34

Figure 28:Pseudo code de la méthode de dichotomie.....	35
Figure 29:Le principe de la méthode de dichotomie	35
Figure 30: L'évolution de R^2 en utilisant la méthode de dichotomie	36
Figure 31: Le pseudo code de l'algorithme de la descente du gradient	37
Figure 32: La convergence de nombre d'estimateurs à l'aide de la descente du gradient.....	37
Figure 33: L'évolution de R^2 à chaque itération de la descente du gradient	38
Figure 34: répartition des données d'apprentissage, validation et d'évaluation	39
Figure 35:architecture du modèle LSTM.....	41
Figure 36: la fonction d'activation Relu.....	41
Figure 37: Plot des résultats du modèle SVR réglé par la méthode de dichotomie.....	43
Figure 38: Résultat du modèle XGBoost réglé par la méthode de la descente du gradient.....	43
Figure 39: Les résultat du modèle LSTM	44
Figure 40: Coefficient d'importance des inputs	45
Figure 41:architecture de la plateforme	47
Figure 42: Les rasters sous format .netCDF	48
Figure 43: diagramme de navigation	49
Figure 44: Le haut de la page d'accueil	52
Figure 45: Page d'acquisition des données de réanalyse.....	53
Figure 46: Bloc des modèles.....	53
Figure 47: importation des données pour créer des modèles d'intelligence artificielle	54
Figure 48 : La page de la personnalisation du modèle SVR.....	55
Figure 49: La page d'entrainement du modèle SVR	56
Figure 50: Page d'entrainement du modèle LSTM	56
Figure 51: La page des résultats du modèle.....	57
Figure 52: le choix du modèle	58
Figure 53: Le débit cumulé observé et simulé à l'aide d'un modèle sauvegardé	58
Figure 54: Le débit observé et simulé en 2011 à l'aide d'un modèle sauvegardé	59
Figure 55: téléchargement des données	59
Figure 56: personnalisation du modèle SRM.....	60
Figure 57: Résultats du modèle SRM	61
Figure 58: La contribution cumulée de la pluie et de la neige pour l'année 2004	61

Liste des tableaux

Tableau 1: répartition des stations de Rheraya par agence	8
Tableau 2: les stations météorologiques de Rheraya et les variables mesurées	9
Tableau 3: Etat d'art de l'utilisation des modèles d'intelligence artificielle pour prédire le débit	13
Tableau 4: Statistiques des précipitations entre 2003 et 2016	19
Tableau 5: Statistiques de la température entre 2003 et 2016	19
Tableau 6: Statistiques de l'humidité au sous bassin de Rheraya entre 2003 et 2016	21
Tableau 7 : statistiques des quatre modèles de SWE entre 2003 et 2016	22
Tableau 8: Les statistiques des valeurs de Snowmelt entre 2003 et 2016	24
Tableau 9: Les statistiques de SCA entre 2003 et 2016.....	25
Tableau 10: Les statistiques du débit au sous bassin de Rheraya entre 2003 et 2016	27
Tableau 11: Comparaison entre grid search et dechotomie	38
Tableau 12 : Hyper paramètres du modèles SVR.....	40
Tableau 13: Hyper paramètres des modèles XGBoost	40
Tableau 14: Hyper paramètres du modèle LSTM.....	41
Tableau 15: Résultats obtenus par les trois modèles	42
Tableau 16: Les résultats des modèles existants.....	46

Introduction générale

L'eau est une matière vitale et la demande sur les ressources en eau s'augmente une année après l'autre, principalement à cause des changements climatiques qui aggravent cette situation, surtout avec la succession des années sèches au Maroc. Malgré qu'il a 149 barrages avec une capacité totale qui peut atteindre 19M m³¹, il a encore besoin d'une politique efficace pour bien gérer ses ressources en eau au vu de la manque des outils opérationnels d'aide à la décision.

C'est dans ce cadre que s'inscrit ce projet de recherche intitulé « prediction de débit au sous bassin de Rheraya à l'aide des modèles d'intelligence artificielle : Recherche et déploiement », qui a pour objectif de créer des modèles d'intelligence artificielle capables de prédire le débit de l'eau dans le sous bassin de Rheraya. Les résultats de cette recherche ont été déployés dans une plateforme considérée comme outil opérationnel pour aider l'agence de bassin hydraulique de Tensift et les autres décideurs à mettre des plans pour bien gérer les ressources en eau du bassin ainsi qu'anticiper les différents événements hydrologiques.

Ce rapport présente le travail de recherche effectué au sein du centre CRSA « Center for Remote Sensing Applications », il est décomposé de cinq chapitres qui sont organisés de la manière suivante :

1. Chapitre 1 : **Contexte général** : dans ce chapitre nous allons mettre notre projet dans son contexte général avec une présentation de la problématique, les objectifs du projet MorSnow et la planification du projet.
2. Chapitre 2 : **Etat de l'art** : ce chapitre présent la zone d'étude et les sources de données qu'on va utiliser ainsi que les modèles d'intelligence artificielle les plus utilisés pour la prédiction de débit.
3. Chapitre 3 : **Analyse des données** : dans ce chapitre on va explorer notre jeu de données, à l'aide des visualisations et des statistiques qui vont nous aider à choisir les variables qu'on va passer à nos modèles.

¹ Statistiques de la ministère de l'équipement de l'eau en 2021