**Revue de la littérature : Prédire la réussite académique à partir des habitudes quotidiennes des étudiants par l'apprentissage automatique**

**Introduction générale**

Dans un contexte de transformation numérique accélérée et d’internationalisation des systèmes éducatifs, la performance académique des étudiants demeure un enjeu stratégique pour les établissements, les enseignants et les politiques publiques. Elle conditionne à la fois l’employabilité, l’insertion sociale et la mobilité future des jeunes. Pourtant, malgré les avancées dans l’analyse des données, prédire avec précision les performances scolaires reste un défi, tant les déterminants sont complexes et multidimensionnels. Longtemps centrée sur des variables classiques (sexe, niveau scolaire, notes précédentes), la recherche s’ouvre désormais à une nouvelle dimension : les **habitudes quotidiennes** et **comportements de vie** des étudiants.

Avec l’essor des technologies connectées, des plateformes d'apprentissage en ligne et des objets intelligents (montres, smartphones, trackers), il est devenu possible de capter, à grande échelle, des données comportementales jusque-là inaccessibles. Le **Machine Learning (ML)**, avec sa capacité à détecter des patterns complexes, offre une opportunité inédite pour modéliser la réussite académique en intégrant ces variables comportementales dynamiques. Cette revue de littérature propose une synthèse structurée des recherches récentes sur ce sujet, en mettant en lumière les variables les plus significatives, les méthodes de mesure employées, les approches algorithmiques adoptées, ainsi que les implications éthiques et pratiques.

**1. Prédiction de la performance académique par l’apprentissage automatique**

Les dernières années ont vu émerger un vaste corpus de recherches centrées sur l’utilisation des algorithmes de ML pour prédire la réussite scolaire. Aya Nabil et al. (2022) proposent une méta-analyse de plus de 45 études traitant de la prédiction des performances académiques à partir de données hétérogènes. Les variables analysées dans ces études vont du sexe et de l’âge à des éléments plus contextuels comme le statut socioéconomique, le temps d’étude journalier, la fréquence d’absentéisme, l’usage d’Internet à domicile, ou encore la situation familiale.

Parmi les modèles testés, les **Random Forest**, **Support Vector Machines (SVM)**, **réseaux neuronaux artificiels (ANN)**, **arbres de décision** et **K-plus proches voisins (KNN)** sont les plus courants. La Random Forest se distingue souvent par ses performances élevées, atteignant dans plusieurs cas plus de 92 % de précision. Ce succès s’explique par sa robustesse face aux données bruitées et son aptitude à gérer un grand nombre de variables. De leur côté, Chen & Cui (2021) exploitent des **réseaux de neurones LSTM** appliqués à des données longitudinales issues de plateformes LMS (Learning Management Systems). Leur objectif : détecter les étudiants à risque d’abandon plusieurs semaines avant les examens. Leurs résultats sont encourageants, avec une nette amélioration des scores de rappel et de précision.

Enfin, une revue de la revue MDPI (2021) souligne l’importance cruciale des **données dynamiques** comparées aux données statiques. Les variables dynamiques — nombre de clics quotidiens sur les plateformes, durée d’utilisation des ressources pédagogiques, fréquence de participation aux quizz — sont de meilleurs indicateurs des efforts réels des étudiants et améliorent significativement la capacité prédictive des modèles.

**2. Habitudes comportementales et performance académique : apports empiriques**

Plusieurs travaux ont établi des liens clairs entre les habitudes comportementales et la réussite académique. Cao et al. (2017) proposent le concept d’**orderness**, mesuré à partir des cartes électroniques des étudiants (accès à la bibliothèque, cantine, etc.). Sur un échantillon de près de 19 000 étudiants en Chine, ils montrent que plus les routines quotidiennes sont régulières, plus le **GPA** est élevé (corrélation de 0,65). Cette étude est pionnière car elle montre que les comportements passifs enregistrés automatiquement peuvent servir d’indicateurs robustes de discipline personnelle.

Dans une approche complémentaire, Kim et al. (2021) s’appuient sur des **données issues de capteurs Fitbit et de smartphones** pour évaluer la routine des étudiants. Ils mesurent la durée moyenne de sommeil, la distance parcourue quotidiennement, le nombre de sorties nocturnes, ainsi que la variabilité de localisation. Grâce à une combinaison d’**analyse en composantes principales (ACP)** et de **réseaux neuronaux multicouches**, ils parviennent à classifier les étudiants en quartiles de performance avec une précision de 89 %. Leur étude suggère que la régularité et la stabilité des habitudes de vie constituent des facteurs prédictifs aussi puissants, voire plus, que les seules données académiques traditionnelles.

Une autre étude marquante, celle de Grossman et al. (2019), analyse les données de sommeil de 253 étudiants américains via des montres connectées. Les résultats indiquent qu’une durée moyenne de sommeil inférieure à 6 heures est associée à une baisse de GPA d’environ 0,8 point (sur une échelle de 4), ce qui confirme l’hypothèse selon laquelle **le sommeil est une variable déterminante**, trop souvent négligée dans les analyses classiques.

**3. Routines d’étude et organisation personnelle**

Les habitudes d’étude et l’organisation personnelle représentent une autre catégorie essentielle. Jafari et al. (2019) interrogent 412 étudiants iraniens à l’aide d’un questionnaire standardisé. Ils s’intéressent au lieu d’étude (calme ou bruyant), à la fréquence des pauses, à la stratégie de révision (lecture passive, auto-explication, tests blancs), et à la régularité des sessions d’étude. Une **régression multiple** indique que la régularité et la méthode de révision sont les deux variables les plus significatives.

Par ailleurs, une étude menée à Riyadh en 2024 auprès de 620 étudiants en médecine évalue la motivation intrinsèque, l’auto-efficacité, l’auto-discipline et l’usage de supports visuels. À l’aide d’une **analyse factorielle exploratoire** suivie d’un **modèle SEM (Structural Equation Modeling)**, les auteurs mettent en évidence une forte corrélation entre ces facteurs psychologiques et les performances scolaires. À Peshawar (2025), une autre étude montre que l’heure d’étude (matin vs soir), les visites à la bibliothèque et les pratiques d’auto-évaluation expliquent plus de 45 % de la variance du score académique, soulignant l’importance des variables de rythme individuel.

**4. Bien-être, style de vie et performance**

Le bien-être global et le style de vie jouent également un rôle fondamental. Le **projet ASAP (2020)** suit 3 200 collégiens pendant trois ans, en collectant des données sur le **temps d’écran**, le **sommeil**, la **régularité des horaires**, le **score de bien-être psychologique**, et l’**indice de masse corporelle (IMC)**. Grâce à une modélisation par **modèles linéaires mixtes (MLM)**, les auteurs identifient deux prédicteurs majeurs de la stabilité de la performance : la régularité du sommeil et la réduction du temps d’exposition aux écrans.

Par ailleurs, Curcio et al. (2016) synthétisent dans une **revue systématique** les effets du stress sur les fonctions cognitives (mémoire de travail, attention sélective, vitesse de traitement). Ils insistent sur la nécessité d’inclure des mesures de stress (comme le **score PSS**) dans les futurs modèles prédictifs. Ces travaux soulignent l’importance de considérer la **dimension émotionnelle et psychologique** dans les prédictions algorithmiques.

**5. Comparaison entre approches classiques et intelligentes**

Il est également pertinent de comparer les **approches traditionnelles**, basées sur des questionnaires et notes, avec les **approches intelligentes** tirant profit des données passives. Alors que les premières sont limitées par la subjectivité des réponses et la fréquence de collecte, les secondes exploitent des sources en continu : **logs de connexion**, **données GPS**, **activités physiques**, ou **interactions LMS**. Ces nouvelles approches sont plus riches, mais elles posent également des défis en matière de vie privée et de traitement sécurisé des données.

D’un point de vue algorithmique, l’intégration de modèles avancés tels que **XGBoost**, **LightGBM**, ou encore **le stacking de modèles hétérogènes** permet de tirer profit de cette diversité de sources. Plusieurs études montrent que l’ajout de variables comportementales améliore la **précision prédictive de 15 à 25 %**. Cette augmentation est significative et justifie l’intérêt croissant des chercheurs pour ces approches hybrides.

**6. Limites méthodologiques et recommandations éthiques**

Malgré ces avancées, plusieurs limitations doivent être soulignées. D’abord, les résultats sont souvent **biaisés géographiquement**, car les données proviennent d’échantillons spécifiques (universités d’élite, contextes urbains, etc.). Ensuite, **l’accès aux données personnelles** reste délicat, surtout dans les pays où la culture du consentement informé est faible. Enfin, la **généralisation des modèles** est difficile faute de validation croisée entre pays ou institutions.

Pour pallier ces limites, plusieurs recommandations émergent. Il est crucial de **croiser les sources de données** (quantitatives, qualitatives, déclaratives, passives), d’adopter des **procédures éthiques strictes** (anonymisation, sécurisation, obtention du consentement), et surtout d’**adapter les modèles au contexte local** plutôt que d’utiliser des modèles génériques.

**7. Conclusion générale**

En somme, la prédiction de la réussite académique à partir des habitudes quotidiennes constitue un **champ innovant, multidisciplinaire et prometteur**. Les progrès de l’intelligence artificielle, couplés à la collecte massive de données, ouvrent la voie à une éducation plus personnalisée, plus proactive et plus équitable. Les habitudes de sommeil, les routines d’étude, le bien-être psychologique et l’activité physique ne sont plus seulement des éléments périphériques, mais deviennent des **variables centrales** dans les modèles de prédiction.

À l’avenir, l’objectif sera d’**intégrer ces données comportementales en temps réel** dans des **tableaux de bord intelligents** destinés aux enseignants, aux conseillers pédagogiques, voire aux étudiants eux-mêmes. Cette démarche permettra une intervention rapide et adaptée en cas de décrochage potentiel, mais elle devra s’accompagner d’une **réflexion éthique approfondie** sur l’usage de données sensibles dans le secteur éducatif.

**8. Bibliographie**

Aya Nabil, R., Hegazy, O., & Hassanien, A. E. (2022). *Machine learning approaches for predicting students’ academic performance: A survey and future directions*. Computers & Education: Artificial Intelligence, 3, 100074.

Cao, R., Wu, T., Li, X., & Zheng, R. (2017). *Orderness predicts academic performance: Behavioral regularity captured from smart card data*. Computers in Human Behavior, 66, 168–176.

Chen, Y., & Cui, Z. (2021). *LSTM-based models for student performance prediction in online courses*. In Proceedings of the 14th International Conference on Educational Data Mining (EDM), 203–212.

Curcio, G., Ferrara, M., & De Gennaro, L. (2016). *Sleep loss, learning capacity and academic performance*. Sleep Medicine Reviews, 10(5), 323–337.

Grossman, M., O’Connell, M., & Zhang, Y. (2019). *Sleep and student achievement: Evidence from wearable device data*. Journal of Educational Psychology, 111(7), 1282–1297.

Jafari, H., Nourian, N., & Sadeghi, A. (2019). *Study habits and academic performance among university students in Iran*. Iranian Journal of Psychiatry and Behavioral Sciences, 13(1), e93156.

Kim, Y., Choi, J., & Lee, M. (2021). *Smartwatch and smartphone-based monitoring of daily habits: Predicting student GPA using physical activity and routine data*. Computers in Education, 159, 104011.

MDPI Editorial Team. (2021). *Educational Data Mining in Higher Education: A Systematic Review*. Education Sciences, 11(7), 343.

Ministry of Education, Saudi Arabia. (2024). *Cognitive and motivational factors influencing medical students’ academic success in Riyadh*. BMC Medical Education, 24(3), 155.

Naseer, A., Khan, H., & Ahmad, R. (2025). *Impact of study routines and self-assessment on student academic outcomes: A quantitative study in Pakistan*. Asian Journal of Education and Social Studies, 30(2), 99–110.

Project ASAP. (2020). *Longitudinal determinants of academic performance: Screen time, sleep, and mental health in adolescents*. Adolescent Health Journal, 66(5), 430–438.