

Modélisation des Mobilités Domicile-Travail

François Bodin*, Laurent Morin*, Alexandre Bettinger*

*prénom.nom@irisa.fr

IRISA, 263 Av. Général Leclerc, 35000 Rennes, France

Résumé. Le transport représente une source majeure d'émissions de gaz à effet de serre. Optimiser les transports en commun et encourager les mobilités douces permet de réduire ces émissions. Pour y parvenir, il est nécessaire de comprendre les schémas décisionnels des usagers en matière de modes de transport. Cela implique la collecte de données personnelles et contextuelles. Le projet BMDec s'intéresse aux mobilités domicile/travail et vise à proposer des outils d'aide à la décision à destination d'acteurs du transport. Ce papier présente deux contributions expérimentales en collecte de données de mobilité ainsi qu'en apprentissage machine pour la prédiction des modes de transport empruntés.

1 Introduction

Le transport représente une source majeure d'émissions de gaz à effet de serre. Optimiser les transports et encourager les mobilités douces permet de réduire ces émissions. Cela demande de comprendre les schémas décisionnels des usagers en matière de modes de transport, en passant par la collecte de données personnelles et contextuelles.

Le projet BMDec s'intéresse aux mobilités urbaines et périurbaines domicile/travail. Il est une initiative pluridisciplinaire réunissant le cabinet SETUR, l'équipe de recherche Logica de l'Irisa, le CREM, UrbanThink ainsi que RatpDev. Le projet est financé par la Région Bretagne et fait partie des activités de la Chaire mobilité dans une ville durable de la Fondation Rennes 1. Ce projet a pour objectifs d'analyser le comportement des usagers pour encourager la diminution des émissions de gaz à effet de serre, et d'intégrer ces comportements au sein d'outils d'aide à la décision. Ainsi, le projet BMDec mise sur la création d'un panel d'usagers interrogés quotidiennement sur les modes de transport qu'ils empruntent.

Le projet comprend deux axes. Le premier concerne la collecte et l'analyse de données relatives aux comportements des usagers. Le second vise à concevoir un système d'aide à la décision reposant sur l'apprentissage machine (ML) afin de générer un modèle des comportements des usagers.

Notre problématique scientifique est donc la suivante : Comment optimiser les transports pour réduire les émissions de gaz à effet de serre en intégrant les comportements des usagers dans des outils d'aide à la décision? Cette problématique mène à la question de recherche suivante : Comment les données contextuelles et personnelles influencent-elles les choix de transport des usagers, et comment utiliser le ML pour modéliser ces comportements et créer des outils pour encourager les modes de transport à faible impact environnemental ?

Le projet se structure autour : (1) d'un panel d'utilisateurs, (2) d'une plateforme d'enquête, (3) de modèles construits par ML, (4) d'un dispositif d'aide à la décision, (5) d'une étude statistique. Ce papier présente deux contributions expérimentales en collecte de données de mobilité ainsi qu'en ML pour la prédiction des modes de transport empruntés par les usagers.

2 État de l'art

L'analyse des mobilités est un domaine en constante évolution. Nos travaux s'intéressent à la collecte des données de mobilité ainsi qu'à la modélisation des déplacements.

2.1 Collecte des données de mobilité

Les progrès dans les technologies de collecte de données ont révolutionné l'analyse des mobilités. Les smartphones, les capteurs dans les véhicules et les systèmes de transport en commun génèrent des données massives qui peuvent être utilisées pour comprendre les schémas de déplacement des individus. Il existe de nombreuses contributions mettant en avant la collecte de données de mobilité dont : (1) La collecte traditionnelle [Audiar \(2018\)](#) qui bien qu'offrant une vision étendue des mouvements, manque souvent de données contextuelles. (2) La collecte basée sur les smartphones [Verzosa et al. \(2017\)](#) ou l'IoT. (3) La collecte issue des foules [Niu et al. \(2020\)](#) qui évalue des plateformes telles que Waze ou Google Maps. (4) La collecte issue des médias sociaux [Nikolaidou et al. \(2018\)](#).

Notons que les collectes de données de mobilité sont souvent incomplètes pour permettre une analyse des comportements car elles n'incluent pas certains types de données comme ceux liés aux profils sociaux économiques des usagers.

2.2 Modélisation des déplacements

La modélisation des déplacements est un domaine vaste visant la prédiction des comportements de mobilité [De Dios Ortúzar et al. \(2011\)](#). Nous notons les principales contributions ou familles de contributions suivantes : (1) Le modèle gravitaire [Erlander et al. \(1990\)](#) qui prédit les flux entre deux endroits en fonction de leur taille et de la distance qui les sépare. (2) La théorie de la demande de transport [Vickrey \(1969\)](#) qui introduit le concept de prix optimal pour le transport. (3) Les modèles de choix discrets [McFadden et al. \(1973\)](#) permettant de prédire les choix individuels en matière de transport en fonction de différentes options. (4) Les modèles d'équilibre dans les transports [Downs \(1962\)](#) qui représentent comment les choix individuels et les interactions entre les agents affectent les flux de transport. (5) Les modèles basés sur les réseaux [Barabási et al. \(1999\)](#) qui s'intéressent à la structure des réseaux de transport pour expliquer la propagation des flux. (6) Le modèle transthéorique [Labarre \(2022\)](#) qui s'intéresse aux changements de comportement en différentes étapes successives. (7) Les Systèmes Multi-Agents (SMA) [Dorri et al. \(2018\)](#) pour les simulations spatiales basées sur des agents. (8) Les modèles basés sur le ML pour analyser les schémas de mobilité et prédire les comportements de déplacement. Ils sont entre autres utilisés pour la prédiction de flux de trafic [Zhang et al. \(2016\)](#), l'optimisation des transports [Mnih et al. \(2015\)](#), la prédiction de la demande de transport en commun [Li et al. \(2019\)](#), la détection d'anomalies dans les données de mobilité [Come et al. \(2021\)](#), la recommandation d'itinéraires [Luca et al. \(2021\)](#).

Notons que le contexte territorial est crucial pour la performance de ces modèles, ce qui souligne l'importance des collectes de données et expérimentations territoriales. Par exemple, la plupart des travaux menés aux États-Unis sont très centrés sur l'usage de la voiture qui est un mode de transport dominant, alors qu'en France il est nécessaire d'étudier une plus grande diversité de modes de transport.

2.3 Positionnement des contributions

Dans le cadre du projet BMDec, nous proposons une démarche de collecte de données traditionnelle s'appuyant sur une plateforme d'enquête dont la teneur et le déploiement constituent des contributions tangibles au domaine des mobilités. Les données issues de cette plateforme ont été enrichies par des données contextuelles provenant de sources externes au projet, afin de répondre à ce manque souvent mentionné dans la littérature. Pour appréhender les évolutions comportementales, nous avons proposé une application aux usagers afin qu'ils puissent y renseigner les modes de transport qu'ils empruntent chaque jour. Nous avons écarté l'option d'une application mobile, jugée trop intrusive et pouvant poser des problèmes d'adoption par les usagers. Les données recueillies représentent plusieurs centaines d'usagers et plusieurs dizaines de milliers de trajets. Nous nous projetons à long terme avec d'autres collectes de données prévues dans d'autres villes.

Concernant la modélisation des déplacements, nos travaux se concentrent sur la prédiction des modes de transport empruntés, en utilisant des algorithmes de ML faisant de la classification multinomiale. Notre contribution en ML est expérimentale dans la mesure où nous avons appliqué des algorithmes de ML existant au contexte spécifique de la prédiction des modes de transport empruntés par les usagers dans le cadre du projet BMDec. Notre contribution présente une approche similaire aux travaux de [Hagenauer et al. \(2017\)](#).

3 Contributions

3.1 Plateforme d'enquête

Notre plateforme de collecte de données est le support d'une méthodologie en plusieurs étapes comportant l'enregistrement d'un panel de volontaires (usagers), la sélection d'un panel représentatif à partir des profils socio-économiques, la collecte de données par l'interrogation journalière des usagers, l'analyse des données et la synthèse. Cinq modules composent le dispositif : (1) le module d'inscription des usagers, (2) le module en charge d'interagir avec les usagers qui produit et synthétise le contexte et les formulaires pour chaque usager, (3) le module de gestion anonymisé des usagers, (4) le module d'analyse et d'agrégation des données, (5) le module de synthèse permettant de produire les données de paiement des usagers.

3.1.1 Phase de sélection

La phase de sélection fait intervenir le module d'inscription des usagers et leur sélection sur plusieurs semaines. Nous proposons quelques figures concernant les répartitions des usagers sélectionnés pour participer à l'étude. La Figure 1 présente la répartition homme/femme par tranche d'âge des usagers. Nous constatons que les usagers sélectionnés représentent un large

Modélisation des Mobilités Domicile-Travail

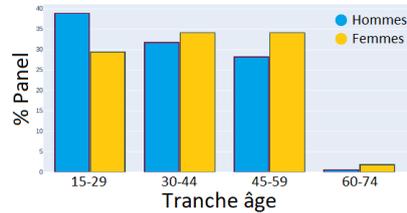
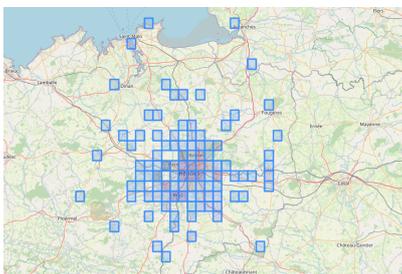
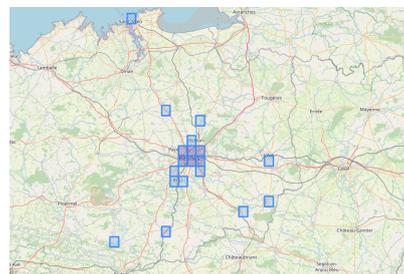


FIG. 1 – Répartition homme/femme par tranche d'âge des usagers



(a) Répartition des domiciles des usagers



(b) Répartition des lieux de travail des usagers

FIG. 2 – Répartition des domiciles et lieux de travail des usagers

spectre d'âges équitablement répartis entre les différentes tranches d'âges mais aussi entre les hommes et les femmes. La Figure 2a présente la répartition des domiciles des usagers. Nous constatons que les usagers sont pour la plupart domiciliés au sein de la zone d'attractivité proche de la métropole rennaise. Quelques usagers sont domiciliés dans des zones plus lointaines. La Figure 2b présente la répartition des lieux de travail des usagers dans l'aire d'attractivité de la métropole rennaise. Nous constatons que les lieux de travail des usagers se situent pour la plupart dans l'aire d'attractivité proche de la métropole rennaise.

3.1.2 Phase de collecte

La phase de collecte dure trois mois et repose sur des sondages personnalisés sous la forme de formulaires accessibles sur internet et notifiés par messagerie et SMS. Elle intègre l'enquête sociologique, l'enquête sur les modalités de transport en voiture et l'interrogation journalière sur la mobilité prévisionnelle pour le lendemain et réelle du jour. Pendant cette phase, les modules de gestion des formulaires et de gestion des usagers sont utilisés pour piloter et monitorer la plateforme. Les formulaires sont produits avec des technologies Web standard et sont conçus pour être lisibles et rapides à remplir. C'est aussi pendant cette phase que deux types de données contextuelles ont été capturés : les opportunités de transport en commun pour chaque trajet en utilisant des services Navitia [Kisio \(2023\)](#) et les caractéristiques des mêmes trajets en voiture.

3.1.3 Phase d'analyse des données

La dernière phase met en œuvre le module d'agrégation et d'analyse des données puis le module de paiement sur la base des données d'assiduité. Des données contextuelles viennent compléter le corpus existant comme des données météorologiques ou des événements susceptibles de perturber les trajets. À l'issue de cette étape, la plateforme fournit pour chaque usager un profil social-économique ainsi que les séries temporelles sur ses choix de mobilités réelles et potentielles. Ces données sont pseudonymisées et dépourvues de coordonnées géographiques.

3.1.4 Sécurisation et conformité au RGPD

Les éléments techniques et les travaux de mise en conformité de la plateforme s'articulent autour de la génération, la gestion et l'analyse des campagnes de sondage. En raison de la nature sensible des données collectées, des principes de sécurisation ont structuré la conception et la réalisation de la plateforme. En effet, l'architecture du système expérimental repose sur des micro-services sécurisés ainsi que sur un confinement des processus de traitement des données. Les données ont été catégorisées en cinq niveaux de sensibilité afin que chaque traitement automatisé soit limité aux accès strictement nécessaires et afin que les accès des opérateurs soient restreints en fonction de leurs responsabilités. Plusieurs dispositifs de sécurité ont ainsi été mis en place, concernant l'intégrité des données, la garantie de disponibilité, la sauvegarde ainsi que la confidentialité des données sensibles. La plateforme ainsi que l'ensemble du processus expérimental ont aussi fait l'objet d'une certification de conformité au RGPD ainsi que de déclarations associées au Délégué à la Protection des Données (DPO). Ces démarches ont été réalisées conjointement par le laboratoire de recherche et le CNRS. À cette occasion, les usagers ont pu avoir accès aux informations relatives au traitement et à l'usage de leurs données.

3.2 Modélisation des comportements du panel

La modélisation du panel vise à prédire les modes de transport empruntés par les usagers dans le but de prendre des décisions dans le cadre de divers scénarios touchant à la logistique, au marketing ou à la décarbonation du trafic. Il s'agit, en ML, d'un problème de classification multinomiale. Nous détaillons cette contribution expérimentale dans la section suivante.

4 Expérimentations

Cette section présente le contexte de nos expérimentations en ML et les résultats associés.

4.1 Jeu de données BMDec

Le jeu de données BMDec comprend des informations sur les profils des usagers et leurs mobilités, pour un panel de plusieurs centaines d'usagers. Au total, il y a environ 27 000 observations de trajets domicile-travail. Les prédicteurs sont donc des informations de profil et de contexte liées aux mobilités et les classes à prédire sont les modes de transport empruntés. Plus précisément, nous avons convergé vers l'utilisation des prédicteurs présentés dans Tab. 1. Ce tableau présente les prédicteurs liés aux profils des usagers ainsi qu'aux contextes des déplacements tout en précisant ceux qui ont été ajoutés (added) aux données d'origine.

Modélisation des Mobilités Domicile-Travail

Predictor	Profile	Context	Added	Description
Gender	✓			User's gender
Age	✓			User's age
Driving	✓			User has a driving license
Handicap	✓			User has reduced mobility
Study	✓			User's study level
Social	✓			User's social situation
Income	✓			User's income level
Work	✓			User's work category
Bicycle	✓			User can use bicycles
T_VS_PT	✓			User prefers train over public transport
PV_VS_PT	✓			User prefers personal vehicle over PT
Destination		✓		Journey is to go work or to go home
Geo_Start		✓		Journey's starting geo position
Geo_End		✓		Journey's ending geo position
Date		✓		Journey's date
Solo		✓		User made the journey alone
Distance		✓	✓	Journey's estimated distance
CO2		✓	✓	Journey's estimated CO ₂ emissions
Alt		✓	✓	Estimated alternative transport modes
Meteo		✓	✓	Journey's meteorological situation

TAB. 1 – BMDec ML - Predictors

Dans le cadre de nos expérimentations, le jeu de données BMDec a été transformé pour jouer sur le volume, le niveau de multinomialité et le bruit. Pour des raisons liées au RGPD, nous ne sommes pas autorisés à partager ce jeu de données. Il pourra cependant être proposé en contribution à l'avenir, sous une forme transformée et en conformité avec le RGPD.

4.2 Algorithmes de la littérature

Nous avons sélectionné différents algorithmes de la littérature : Naive Bayes (NBayes), Arbre de décision (DTree), Régression logistique (LogReg), Réseau de neurones (NN), K Plus Proches Voisins (KNN), Machine à vecteur de support (SVM). Tous les algorithmes sont exécutés plusieurs fois, c'est-à-dire qu'ils créent leur modèle plusieurs fois dans des conditions variables en fonction de la façon dont les données sont réparties entre les données d'entraînement et les données de test. Chaque exécution est évaluée à l'aide de critères d'évaluation dont la moyenne est ensuite calculée pour toutes les exécutions. Enfin, il convient de noter que les algorithmes sont exécutés de manière normale, sans utiliser le boosting par exemple.

4.3 Critères d'évaluation

La littérature et les outils de développement proposent de nombreux critères d'évaluation des algorithmes de ML. Nous avons choisi d'utiliser des critères d'évaluation assimilés à des scores : Accuracy (Acc), Precision (Prec), F1, FBeta (Fb), Recall (Rec), Jaccard (Jac), Matthews CC (Mat), Hamming Loss (Ham). Notons que plus le score est élevé, meilleur il est, exception faite de l'Hamming Loss qui fonctionne à l'opposé.

4.4 Résultats

Les résultats de nos expérimentations sont présentés dans Tab. 2.

Les résultats montrent que DTree et KNN surpassent les autres algorithmes de la littérature et semblent donc bien adaptés au jeu de données BMDec dans le cadre d'une classification multinomiale. D'autres expérimentations sont nécessaires en utilisant différentes variations du

	Acc	Prec	F1	Fb	Rec	Jac	Mat	Ham
NBayes	0,75	0,69	0,69	0,69	0,7	0,57	0,7	0,25
DTree	0,92	0,89	0,89	0,89	0,89	0,84	0,9	0,08
LogReg	0,68	0,61	0,59	0,6	0,6	0,46	0,58	0,32
NN	0,83	0,76	0,75	0,76	0,76	0,66	0,79	0,17
KNN	0,91	0,86	0,86	0,86	0,86	0,8	0,89	0,09
SVM	0,68	0,52	0,51	0,51	0,54	0,4	0,58	0,32

TAB. 2 – BMDec ML - Scoring Results

jeu de données BMDec mais aussi d’autres algorithmes de la littérature. Cela nous permettrait de consolider les résultats obtenus avant de déployer la solution prédictive définitive.

5 Conclusion

Dans le cadre du projet BMDec nous nous demandons comment optimiser les transports en intégrant les comportements des usagers dans des outils d’aide à la décision. Cette problématique nous mène à nous demander comment les données contextuelles et personnelles influencent les choix de transport des usagers, et comment utiliser le ML pour modéliser ces comportements et créer des outils d’aide à la décision.

Pour répondre à ces questions, nous avons proposé une plateforme pour la collecte de données de mobilité auprès d’un panel d’usagers suivis de manière quotidienne sur plusieurs mois. Nous avons enrichi les données collectées avec des données contextuelles annexes et nous avons utilisé l’ensemble des données pour modéliser le comportement des usagers à l’aide du ML. Les premiers résultats mettent en lumière la capacité de certains algorithmes de la littérature à prédire les modes de transport empruntés par les usagers avec une précision dépassant les 90%. Nos contributions sont donc expérimentales et pourront servir de référence aux domaines de la collecte et de l’analyse des données de mobilité. Nous envisageons d’expérimenter avec d’autres modèles pour affiner les résultats actuels.

Nos résultats ouvrent la voie à des applications potentielles des modèles pour des tâches prédictives permettant l’analyse de scénarios de déplacements, la prise de décisions, l’optimisation logistique et marketing de l’offre de transport, la prise en compte de la multi-modalité ou encore l’appréciation de l’empreinte carbone dans le but de la réduire. Ces applications dépendent des principaux acteurs du projet.

Par ailleurs, il faudra déterminer les variables contextuelles et personnelles ayant le plus d’impact sur la précision des modèles.

Enfin, nous validerons notre approche en la confrontant à d’autres groupes démographiques. À ce titre, une nouvelle collecte de données est prévue en 2024 au sein de Brest Métropole.

Références

Audiar (2018). Enquête ménage-déplacements.

- Barabási, A.-L. et al. (1999). Emergence of scaling in random networks. *science* 286(5439), 509–512.
- Come, E. et al. (2021). Machine learning and data mining for urban mobility intelligence. In *ESANN 2021*, pp. pp–453.
- De Dios Ortúzar, J. et al. (2011). *Modelling transport*. John Wiley & sons.
- Dorri, A. et al. (2018). Multi-agent systems : A survey. *IEEE Access* 6, 28573–28593.
- Downs, A. (1962). The law of peak-hour expressway congestion. *Traffic Quarterly* 16(3), 393–409.
- Erlander, S. et al. (1990). *The gravity model in transportation analysis : theory and extensions*, Volume 3. Vsp.
- Hagenauer, J. et al. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications* 78, 273–282.
- Kisio (2023). Navitia.io : L’api ouverte de référence dédiée à la mobilité.
- Labarre, F. (2022). *Favoriser le report modal par le management de la mobilité*. Theses, École Nationale des Travaux Publics de l’État [ENTPE].
- Li, Y. et al. (2019). Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering* 32(6), 1079–1091.
- Luca, M. et al. (2021). A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* 55(1), 1–44.
- McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.
- Mnih, V. et al. (2015). Human-level control through deep reinforcement learning. *nature* 518(7540), 529–533.
- Nikolaidou, A. et al. (2018). Utilizing social media in transport planning and public transit quality : Survey of literature. *Journal of Transportation Engineering* 144(4), 04018007.
- Niu, H. et al. (2020). Crowdsourced data mining for urban activity : Review of data sources, applications, and methods. *Journal of Urban Planning and Development* 146(2), 04020007.
- Verzosa, N. et al. (2017). Smartphone-based travel surveys : a review.
- Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review* 59(2), 251–260.
- Zhang, J., Y. Zheng, D. Qi, R. Li, et X. Yi (2016). Dnn-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 1–4.

Summary

Optimizing public transport and encouraging soft mobility can help reduce greenhouse gas emissions. It implies to understand users’ decision-making patterns regarding transports. The BMDec project focuses on home/work mobility, and aims to provide decision-support tools for transport stakeholders. This paper presents two experimental contributions in mobility data collection and machine learning for the prediction of transport modes used.