

Présentation Projet IA :

Prédiction du taux de churn pour la compagnie de Telecom Datalabel

Soutenance du module RNCP36129 BC02
par Thomas Zilliox

- 1/ Présentation du projet : Problématique, présentation de la solution et composition de l'équipe
- 2/ Cahier des charges : Activités à planifier, Risques possibles du projet et Proposition de solution
- 3/ Estimation du budget pour le projet
- 4/ Composition de l'équipe en charge du projet et leurs missions
- 5/ Sélection des prestataires externes
- 6/ Mise en place de la formation : former les équipes avant et pendant le projet
- 7/ Préparation des supports de formation

1) Présentation du projet



Problématique :

Notre client est une compagnie de Telecom nommée Datalabel, basée en Europe, a du mal à fidéliser ses consommateurs. Notre principal but est de savoir ce qui serait le plus bénéfique pour notre entreprise financièrement : fidéliser les clients ou en acquérir de nouveau ? Comment éviter que les clients ne changent de compagnie de Telecom ?



Présentation de la solution :

- Création d'un algorithme de prédiction basé sur la classification pour savoir si les nouveaux clients risquent de partir de Datalabel, algorithme basé sur les features déjà disponibles dans le CRM de Datalabel, en fonction des caractéristiques des clients et leur historique d'abonnement.
- Visualisation des résultats obtenus à l'aide d'un dashboard



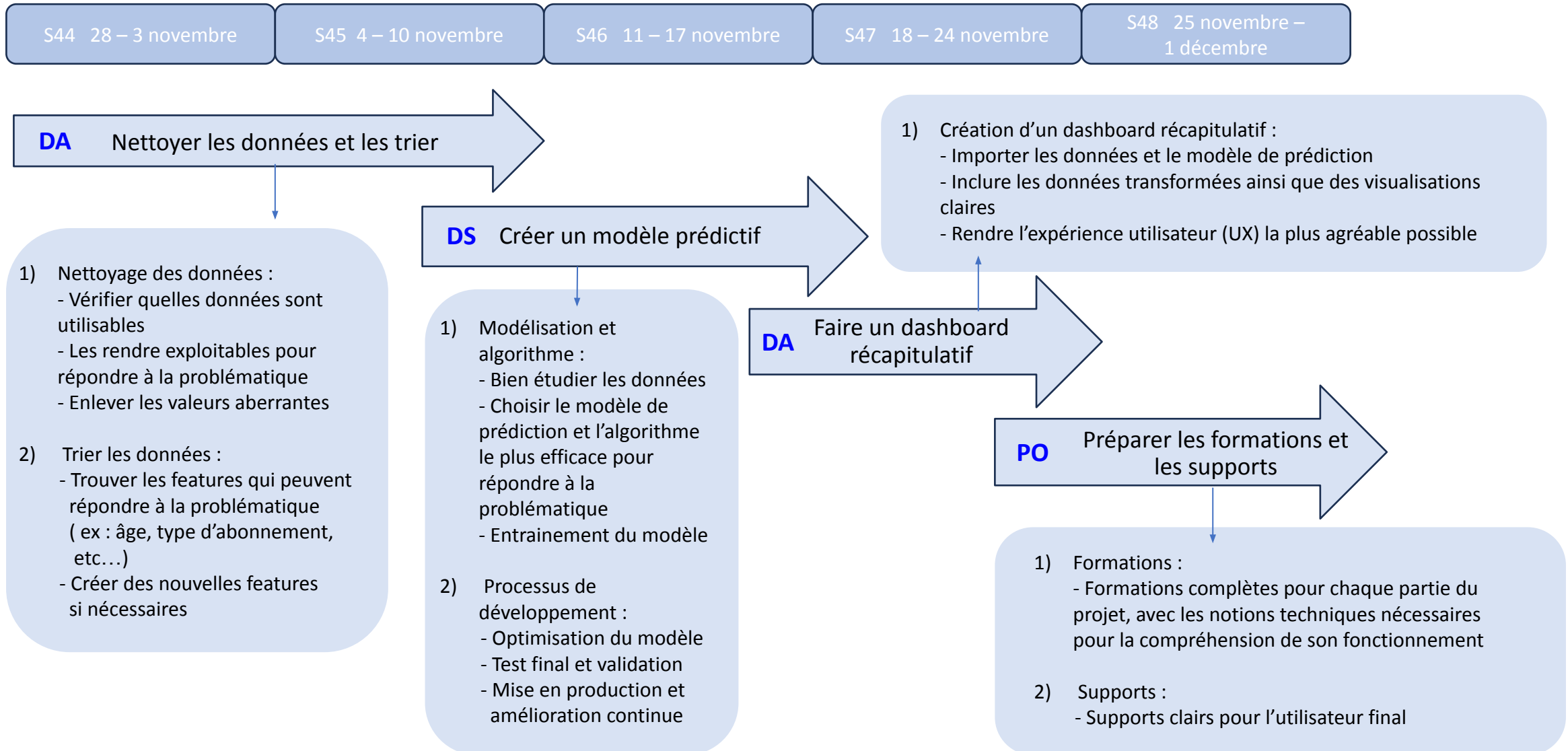
Composition de l'équipe :

- 1 Data Analyst, possédant 3 à 5 années d'expérience
- 1 Data Scientist possédant 3 à 5 années d'expérience
- 1 Product Owner possédant 3 à 5 années d'expérience
- 1 Data Engineer externe possédant 3 à 5 années d'expérience



Objectif Date de lancement : Un mois après le début du projet, lancement du Minimum Viable Product (25 novembre 2024).

2) Elaboration du cahier des charges : Activités à planifier

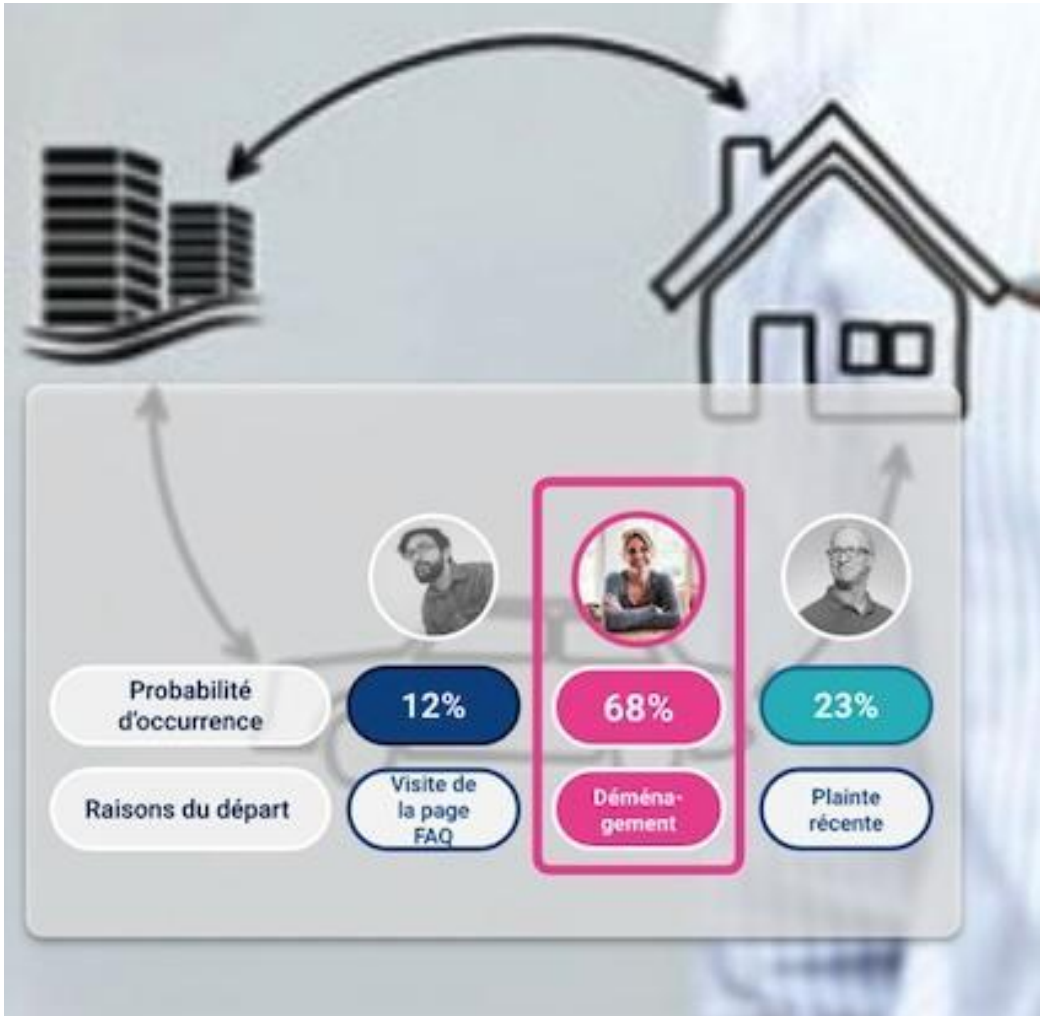


2) Elaboration du cahier des charges : Risques du projet et évaluation de leur gravité

Étapes du Projet	Risques	Criticité	Détails des risques
Nettoyer les données et les trier	1) Mauvaise Qualité des données	Cruciale	<ul style="list-style-type: none"> Données manquantes/incomplètes Données non normalisées Données obsolètes
	2) Gouvernance des données	Cruciale	<ul style="list-style-type: none"> Problèmes de conformité (RGPD) Accès non autorisé Manque de traçabilité des données
Créer un modèle prédictif	1) Problème de modélisation et d'algorithme	Cruciale	<ul style="list-style-type: none"> Modèles biaisés, explicabilité limitée en fonction du choix du modèle de prédiction (exemple : random forest) Performances insuffisantes Dépendance excessive à l'algorithme
	2) Processus de développement	Élevée	<ul style="list-style-type: none"> Mauvaise priorisation Dérive des objectifs Manque de communication entre les équipes
Faire un dashboard récapitulatif	1) Risques liés à l'expérience utilisateur	Moyen	<ul style="list-style-type: none"> Mauvaise adoption du produit Manque de confiance dans les résultats Problème d'expérience utilisateur (UX)
Préparer formations et supports	1) Problèmes de formations	Moyen	<ul style="list-style-type: none"> Formations incomplètes ou pas assez claires
	2) Problèmes de supports	Moyen	<ul style="list-style-type: none"> Supports incomplets ou pas assez clairs



Solution trouvée pour détecter les churners



Nous vous proposons un algorithme de machine learning nommé **Random Forest**.

Objectif : détecter les risques d'attrition en temps réel.

Type d'algorithme : Supervisé / Classification



Pour contre-balancer la non-explicabilité de l'algorithme de classification Randomforest, nous vous proposons également d'utiliser un autre algorithme nommé **Word2vec**.

Objectif : donner aux employés des éléments décisifs pour fidéliser les clients (les motifs de départs)

Type d'algorithme : Semi supervisé



Pourquoi avoir choisi ces deux algorithmes ?

Algorithme de classification supervisé **RandomForest** :



1. **Algorithme supervisé** : Entraîne un modèle à partir de données étiquetées (ex : clients ayant churné ou non) pour faire des prédictions sur de nouvelles données.
2. **Algorithme de classification** : Prédit une catégorie ou classe, dans ce cas "churn" ou "pas churn", parfaitement adapté au problème du churn.
3. **Random Forest** : Combinaison de plusieurs arbres de décision qui améliore la robustesse et la précision des prédictions, tout en étant flexible et performant pour des problématiques complexes comme la prédiction du churn des clients.

Algorithme semisupervisé **Word2vec** :



1. **Approche semi-supervisée** : mélanger un ensemble de données non labelisées et labelisées, dans un contexte où les interactions et les données clients sont souvent peu ou mal annotées.
2. **Word2vec** : Capturer des relations sémantiques, rapide et **scalable** sur de grands volumes de données, ce qui est crucial pour analyser les interactions avec les clients de Datalabel chaque jour. Cela permet d'analyser rapidement ces données en temps réel et de fournir des informations aux employés pour qu'ils puissent prendre des décisions immédiates.



1. Amélioration de l'expérience client (Customer Experience)

- Optimisation du service client
- Simplification des processus



2. Offres personnalisées et segmentées

- Segmentation des clients
- Offres de rétention personnalisées
- Cross-selling et up-selling.



3. Programmes de fidélisation et de récompenses

- Programme de fidélité
- Inciter les clients à renouveler leur abonnement



4. Innovation produit et différenciation

- Nouveaux services digitaux
- IoT et solutions smart home



3) Elaboration du budget du projet



	Nombre	Par jour	Par Semaine	Nombre de jours de travail estimé	Total
Budget Ressources humaines interne					
=> Data Analyst	1	205 euros	1025 euros	17	3485 euros
=> Data Scientist	1	250 euros	1250 euros	8	2000 euros
Budget Ressources humaines externe					
=> Data Engineer	1	620 euros	3100 euros	3	Variable en fonction du besoin
Budget Outils					
=> Management : Jira	1	0	0	0	0
=> Base de données et stockage	1				2,30 euros/mois
Budget imprévu fourchette haute (15 à 30% du budget total)		78,75/157,50 euros	425,15/850,50 euros		517,50/1500 euros
Total du budget		110/208 euros	885,20/1512 euros	20 journées	6000/6695 euros

*Chaque chiffre représente une moyenne en fonction des informations recueillies sur Internet

**Pour chaque salaire moyen, cela correspond au salaire moyen de personnes avec 3 à 5 années d'expérience



Composition de l'équipe en charge du projet :

Membre de l'équipe	Nombre	Type de prestataire	Nombre d'années d'expérience	Missions
Data Analyst	1	Interne	3 à 5 années	<ul style="list-style-type: none"> Récolter les données Nettoyer et trier les données Analyser les données Créer un dashboard Power BI contenant les résultats du modèle de prédiction
Data Scientist	1	Interne	3 à 5 années	<ul style="list-style-type: none"> Créer un modèle prédictif : <ul style="list-style-type: none"> - Trouver l'algorithme le plus efficace - Améliorer la performance de l'algorithme au fur et à mesure du temps
Data Engineer	1	Externe	3 à 5 années	<ul style="list-style-type: none"> S'assurer de la fiabilité ainsi que de la disponibilité du pipeline de données Résoudre les problèmes techniques qui pourraient survenir avec le temps
Product Owner	1	Interne	3 à 5 années	<ul style="list-style-type: none"> Manager l'équipe et s'assurer du bon fonctionnement du projet et d'une bonne communication entre chaque membre de l'équipe Prendre rendez-vous fréquemment avec le client pour s'assurer que le projet répond bien à ses attentes Garder un œil sur la marge brute du projet ainsi que du respect des deadlines

5) Sélection des prestataires

Besoins immédiats :

- Assurer la fiabilité et la disponibilité du pipeline de données
- Garantir le bon fonctionnement des flux de données et intervenir rapidement en cas de problème

Exemples concrets de problèmes pouvant nécessiter un Data Engineer :

- Besoin de mettre à jour en streaming les données
- Assurer l'accès au dashboard
- Se connecter à des applications partenaires pour enrichir la base d'entraînement : pour l'instant , le Data Scientist n'utilise que les données de Datalabel, mais on pourrait mettre en place une API qui récolterait également les données par l'application mobile de Datalabel

6) Piloter la mise en place de la formation pour l'équipe



1. Former les équipes avant le projet : préparation et alignement

- **Connaissance des fondamentaux du projet** : Réaliser une première réunion expliquant à toute l'équipe le déroulement global du projet
- **Formation sur le domaine métier** : S'assurer que chaque membre de l'équipe connaisse les rôles de chacun des autres membres de l'équipe, s'acculture aux notions de Machine Learning pour comprendre les algorithmes utilisés et comprenne le processus de prédiction du début jusqu'à la fin
- **Respecter la protection des données** : Bien définir les règles de protection des données (RGPD) ainsi que de l'IA Act

2. Former les équipes pendant le projet :

- **Réaliser une formation par semaine** d'une heure pour former les autres membres de l'équipe au fonctionnement de la tâche de la semaine actuelle (Powerpoint)
- **Un feedback réalisé à la fin de chaque semaine** pour mettre au clair le travail qui a été effectué de manière précise (Google Docs)
- **Une plateforme pour poser des questions** par rapport à l'incompréhension du fonctionnement du projet (Canal Slack spécifique)
- **Maintenir une collaboration entre les équipes** pour interpréter les résultats du modèle

6) Piloter la mise en place de la formation pour l'équipe

Premier jour de la semaine

Formation de une heure par rapport à ce qui va être fait durant la semaine.

Durant la semaine

Questions et réponses disponibles à tout moment sur un canal de la plateforme Slack.

Dernier jour de la semaine

Feedbacks de chaque équipe par rapport à ce qui a été fait durant la semaine.

Outils
utilisés :





Transmissions des présentations de formation

- Transmissions des documents Powerpoint utilisés lors des formations pour notre équipe afin de fournir aux utilisateurs les détails du fonctionnement du Projet IA.

Introduction au projet

- Détail des objectifs mesurables du projet
- Explications des normes RGPD et de l'IA Act
- Explications du contexte métier

Concepts de base en IA

- Machine Learning
- Algorithme de Classification
- Notions générales de Data Science

Notions de churn

- Sensibilisation de l'importance de baisser le churn de l'entreprise
- Essayer de quantifier les bénéfices possibles apportés par l'algorithme



Il faut toujours être **vigilant aux différents e-mails** que les utilisateurs pourraient nous envoyer en fonction de leurs interrogations, mais aussi par rapport aux potentiels problèmes techniques rencontrés lors de l'utilisation du dashboard final.



Une solution optimale serait de mettre en place **un espace Q&A** pour que les clients puissent poser leurs interrogations de manière efficace à notre équipe de formation et notre support.

Merci pour votre attention.