

Общая документация проекта
«Оптимизация процесса
легирования сталей»

Содержание

Введение	3
Термины и определения	3
Цель проекта	4
Этапы выполнения проекта	5
Исследовательский анализ данных	5
Построение модели предсказания химического состава шлака	7
Построение модели для оптимизации расхода извести	8
Перспективы развития проекта	9
Используемые технологии	9

Введение

Легирование — один из этапов обработки стали. На данном этапе расплавленная сталь заливается в специальный металлический ковш, изнутри облицованный кирпичом, и греется электричеством, чтобы получить нужную химию сплава. В процессе выводится сера, регулируются легирующие присадки, процентовка железа, окислы алюминия, кремния.

На данном этапе осуществляется следующая последовательность действий:

1. Перед началом нагрева берутся первые две пробы: химический состав стали и химический состав шлака.
2. Затем вносятся необходимые добавки, повышается температура и состав продувается газом, в результате чего шлак реагирует и всплывает.
3. Затем сталь перемешивают и проводят вторые измерения химического состава стали и шлака.

Если в результате получен необходимый химический состав — начинается следующий этап технического процесса.

Термины и определения

Дисперсия — мера разброса данных относительно среднего значения.

Выбросы — экстремальные значения, которые отличаются от других наблюдений данных. Они могут указывать на изменчивость измерения, экспериментальные ошибки или новизну.

Категориальный признак — признак, значения которого обозначают принадлежность объекта к какой-то категории.

Вещественный признак — признак, значения которого являются вещественными числами.

Целевой признак — значение, которое необходимо предсказывать в задаче машинного обучения.

Независимый признак — значение, на основе которого делается предсказание в задаче машинного обучения.

Цель проекта

Цель проекта — проанализировать данные с производства, чтобы построить модель. Эта модель по данным, полученным после первого измерения химического состава стали и шлака, позволит прогнозировать химический состав шлака в конце этапа легирования стали.

Дополнительной задачей является исследование возможностей оптимизации расхода извести (одной из вносимых в процессе легирования стали добавок) на производстве с сохранением целевого химического состава шлака.

Этапы выполнения проекта

Проект разделен на три последовательных этапа — исследовательский анализ данных, построение модели предсказания химического состава шлака, построение модели для оптимизации расхода извести.

Исследовательский анализ данных

Основой для построения любой модели машинного обучения являются данные. От их качества и количества, а также от методов последующей их обработки напрямую зависит будущая точность модели.

На данном этапе производится анализ и делаются выводы по следующим характеристикам данных:

- наличие и количество отсутствующих значений;
- нули в данных;
- дисперсия;
- выбросы в данных.

Исходные данные представлены в виде таблицы и имеют следующую размерность:

- 7041 объект;
- 84 признака у объекта.

Главной задачей проекта является построение модели предсказания химического состава шлака после второго измерения, поэтому особенно важно исследовать качество предоставленных данных по целевым признакам.

Целевой признак	Кол-во проп.	Мин.	Макс.	Среднее значение	Медиана	Мода	Станд. откл-е
химшлак последний Al ₂ O ₃	2299	2.4	14.2	4.6	4.5	4.4	0.96
химшлак последний CaO	597	35.9	67.5	56.4	56.8	57.6	3.78
химшлак последний FeO	597	0.1	4.2	0.6	0.6	0.5	0.22
химшлак последний MgO	598	0.3	32.4	9.0	8.9	8.7	3.16
химшлак последний MnO	597	0.04	1.45	0.12	0.11	0.08	0.06
химшлак последний R	597	1.6	3.1	2.4	2.3	2.2	0.22
химшлак последний SiO ₂	2299	16.4	28.7	24.0	24.4	26.0	2.19

Таблица характеристик целевых признаков

Остальные 77 признаков являются независимыми — по ним будет строиться прогноз значений целевых признаков. В результате их анализа были сделаны следующие выводы:

- В исходных данных есть 10 признаков и 598 объектов, у которых количество отсутствующих значений превышает порог 33% – эти признаки и объекты не следует использовать при обучении модели.
- Нулей в данных нет.
- Некоторые из признаков имеют очень низкую дисперсию, что говорит об их малой информативности.
- Данные содержат выбросы, которые для более корректной работы алгоритма машинного обучения следует удалить.

Построение модели предсказания химического состава шлака

Основной задачей проекта является построение модели машинного обучения, которая смогла бы с достаточной точностью предсказывать химический состав шлака в конце процесса легирования по данным после первого измерения.

В ходе исследования командой были изучены различные подходы к обработке данных и построению алгоритма машинного обучения, в результате удалось достичь достаточно высокой точности предсказаний для четырех из семи целевых признаков. Метрикой качества предсказаний была выбрана MAPE (mean absolute percentage error), которая показывает среднее отклонение предсказания модели от истинного значения, выраженное в процентах.

Также был предложен альтернативный подход к построению модели, когда предсказания для целевых признаков делаются последовательно, и каждая следующая модель использует для предсказания результаты работы предыдущей. Такой подход показал более высокую точность, однако требует дальнейшего изучения.

По результатам можно сделать следующие выводы:

- 4 целевых признака (Al_2O_3 , CaO , R и SiO_2) предсказываются с достаточно высокой точностью.
- Оставшиеся 3 целевых признака имеют достаточно высокую ошибку предсказания — имеющихся данных недостаточно для обеспечения необходимой точности.

Значения метрики MAPE для обоих подходов указаны в таблице:

Цел. призн.	химшлак последний Al ₂ O ₃	химшлак последний CaO	химшлак последний FeO	химшлак последний MgO	химшлак последний MnO	химшлак последний R	химшлак последний SiO ₂
I подход	9,9%	4,7%	28,5%	19,2%	27%	5,6%	6,6%
II подход	2,68%	0,75%	6,50%	5,51%	7,15%	0,40%	0,76%

Построение модели для оптимизации расхода извести

На данном этапе проекта необходимо построить модель предсказания расхода извести на основе тех же исходных данных. С помощью данной модели необходимо проанализировать границы возможных значений расхода извести, при которых сохраняется целевой химический состав шлака. Найденная нижняя граница будет являться оптимальным количеством извести, которое нужно использовать на производстве, что позволит сократить расходы.

Для решения данной задачи был предложен подход, заключающийся в разбиении непрерывного целевого признака (расход извести) на три интервала. Интервалы были подобраны так, чтобы обеспечить баланс классов — каждому интервалу должно принадлежать одинаковое количество объектов. Такой же подход был применен и ко всем независимым признакам. Каждому интервалу была дана метка класса “0”, “1” или “2”. Значения всех признаков были заменены на соответствующие им метки интервалов. Таким образом, в результате была получена таблица, которая вместо непрерывных значений содержит в себе только метки классов. По таким данным была обучена модель классификации.

Метрика ассигасы (точность предсказаний, выраженная в процентах) данной модели равна 47%. На текущий момент это недостаточное значение, поэтому работу над этой задачей следует продолжить.

Перспективы развития проекта

На основе построенной модели предсказания химического состава шлака может быть создан цифровой подсказчик — программное обеспечение, которое сможет предсказывать количество необходимых добавок, чтобы получить целевой химический состав шлака.

Имеющиеся результаты по разработке модели, оптимизирующей расход извести можно использовать для дальнейших исследований в этом направлении.

Используемые технологии

В ходе выполнения проекта использовался язык программирования Python и следующие библиотеки:

- Pandas;
- Matplotlib;
- Numpy;
- Scikit-Learn.