# **Augmenting Agronomic Data for Predictive Modeling: A Feature Engineering Report for Fertilizer Optimization**

## **Executive Summary**

**Objective:** The primary objective of this report is to detail the transformation of a foundational dataset, Fertilizer.csv, into a feature-rich, machine-learning-ready format specifically optimized for training a sophisticated gradient boosting model such as XGBoost. The initial dataset, while providing basic N-P-K values, lacks the granularity required to model the complex, non-linear dynamics of fertilizer performance in real-world agricultural systems.

**Methodology:** A systematic, expert-driven analysis was conducted on an extensive corpus of over 70 documents, including technical data sheets, agricultural extension articles, and scientific publications. This process involved the extraction, synthesis, and structuring of previously unstructured information to engineer a suite of new, high-value predictive features. The methodology focused on quantifying qualitative descriptions, inferring causal relationships between fertilizer properties and agronomic outcomes, and codifying domain-specific knowledge into a machine-readable format.

**Key Augmentations:** The original dataset has been substantially enhanced with the addition of over 20 new features, meticulously categorized into three critical domains to provide a holistic view of each fertilizer product:

1. **Physicochemical Properties:** These features define the intrinsic chemical and physical nature of the fertilizer, including Physical\_Form, Solubility\_Class, granular size specifications, and a detailed breakdown of nitrogen forms (N\_Form\_Urea\_pct, N\_Form\_Ammoniacal\_pct).
2. **Agronomic & Application Characteristics:** This category describes how the fertilizer behaves within an agricultural context, featuring crucial attributes such as Release\_Rate, recommended application timing and methods, and crop-specific usage protocols.
3. **Environmental & Safety Profiles:** These features quantify the potential impacts of the fertilizer on the growing environment, including its effect on soil pH (Soil\_pH\_Effect), its propensity for nutrient loss (Leaching\_Potential), and its risk of causing damage to young plants (Phytotoxicity\_Risk).

**Anticipated Outcome:** The resulting augmented dataset, Fertilizer\_Augmented.csv, is designed to empower an XGBoost model to move beyond simple correlations and achieve a significantly higher level of predictive accuracy. By providing features that represent the underlying mechanisms of nutrient delivery, soil interaction, and environmental risk, the model will be capable of learning the complex, high-order interactions that govern fertilizer efficacy. This will enable more precise, data-driven recommendations for fertilizer selection, ultimately leading to improved crop yields, enhanced nutrient use efficiency, and more sustainable agricultural practices.

## **The Imperative for Advanced Feature Engineering in Agronomic Modeling**

The predictive power of any machine learning model is fundamentally constrained by the quality and richness of the data it is trained on. While algorithms like XGBoost are exceptionally powerful, their ability to deliver accurate and actionable insights in a complex domain like agriculture is contingent upon a feature set that captures the nuances of the system being modeled.

### **Analysis of the Original Dataset's Limitations**

The initial Fertilizer.csv dataset provides a foundational but ultimately superficial view of each product.1 It contains the name, the percentage of the three primary macronutrients (Nitrogen, Phosphorus, Potassium), and a simple list of suitable soils and crops. An analysis of this structure reveals critical limitations for advanced modeling:

* **Lack of Mechanistic Detail:** The dataset describes *what* is in the fertilizer (N-P-K percentages) but provides no information on *how* it works. For example, the model cannot distinguish between two fertilizers both labeled '28-28-0'. One might be a fast-acting, water-soluble powder intended for fertigation, while another could be a coated, controlled-release granule designed for season-long feeding.2 Despite having identical NPK ratios, their agronomic behavior, application timing, and environmental impact are vastly different. The original dataset renders these two products indistinguishable.
* **Inability to Generalize:** The model can learn correlations (e.g., 'DAP' is associated with 'Wheat') but cannot understand the underlying reasons for this association. It does not know that DAP provides ammoniacal nitrogen and a high concentration of phosphorus, which are critical for the early root development and tillering of wheat.4 Without this mechanistic understanding, the model cannot generalize its knowledge to new, unseen fertilizers or predict performance under varying environmental conditions.
* **Oversimplification of Complex Interactions:** Crop yield is not a simple function of N, P, and K. It is the result of intricate interactions between the fertilizer's chemical form, the soil's pH, the soil's texture, water availability, and the crop's specific growth stage. The original dataset fails to provide the necessary features to model these interactions.

### **Unlocking XGBoost's Potential with Granular Features**

XGBoost and other tree-based ensemble models are particularly adept at discovering complex, non-linear relationships and high-order interactions within data. However, they can only find interactions between the features they are given. The process of feature engineering, therefore, is not merely about adding more data; it is about providing the model with the fundamental building blocks of domain knowledge.

By engineering features that represent core agronomic concepts, we allow the model to learn sophisticated rules. For instance, a well-featured model could deduce that the agronomic value of a high Phosphorous\_Percentage is maximized only when the Soil\_pH\_Effect is not 'Acidifying' (as low pH can limit phosphorus availability) and the Leaching\_Potential of the accompanying nitrogen is 'Low' (preventing loss of the other key nutrients).6 This level of inference is impossible with the original dataset.

### **Introduction to Feature Categories**

To address these limitations, a comprehensive feature engineering process was undertaken. The new features are logically grouped into categories that reflect a multi-faceted understanding of fertilizer products, moving from their intrinsic properties to their real-world behavior and impact.

1. **Physicochemical Features:** These are the intrinsic, measurable properties of the fertilizer itself. They are the "what it is" features that define its fundamental composition and form, such as its physical state (granular, liquid), solubility, and the specific chemical forms of its nutrients.
2. **Agronomic and Environmental Features:** These features describe how the fertilizer behaves and interacts within an agricultural ecosystem. They are the "how it works" and "what it does" features, encompassing its nutrient release rate, its effect on soil chemistry, and its potential risks to the environment and the crop itself.
3. **Crop-Specific Application Features:** These are tactical features that provide detailed guidance on the optimal use of a fertilizer for specific crops. They capture critical information on application timing, methods, and ideal environmental conditions, enabling the development of highly specialized, crop-centric predictive models.

## **Augmentation with Physicochemical Features**

The first step in enriching the dataset is to codify the fundamental physical and chemical properties of each fertilizer. These features serve as the foundation upon which more complex agronomic behaviors are built.

### **3.1. Physical State and Formulation**

The physical form of a fertilizer is a primary determinant of its handling, application method, and rate of nutrient release.

* **New Features:** Physical\_Form (Categorical), Solubility\_Class (Categorical), Is\_Organic (Boolean)
* **Rationale and Data:** The physical form dictates the mode of application. **Granular** fertilizers are typically applied via broadcasting or banding with mechanical spreaders and are common for basal applications.8  
  **Powder**, **Liquid**, and **Gel** formulations are generally designed to be water-soluble, making them suitable for application through irrigation systems (fertigation) or as foliar sprays directly onto the plant leaves.3 The distinction between  
  **organic** and synthetic fertilizers is also critical; organic sources typically release nutrients more slowly as they require microbial breakdown, which also contributes to long-term soil health, whereas synthetic fertilizers provide rapid, targeted nutrition.11

A critical relationship emerges when considering the physical form of a fertilizer in conjunction with its solubility. Products formulated as liquids, gels, or fine powders are, by design, highly water-soluble.3 This high solubility facilitates rapid nutrient release, which is beneficial for correcting acute nutrient deficiencies or feeding crops during critical growth stages.14 However, this same property introduces a significant environmental risk. For mobile nutrients like nitrogen (particularly in its nitrate form), a direct causal link can be established between a liquid or powder physical form, high solubility, and an elevated potential for environmental loss through leaching if the nutrient release is not perfectly synchronized with crop uptake.18 By creating distinct features for

Physical\_Form and Solubility\_Class, we provide the model with the inputs needed to learn this crucial trade-off between nutrient availability and environmental risk.

### **3.2. Granulometry**

For solid fertilizers, the size and uniformity of the granules are important quality and performance indicators.

* **New Features:** Granule\_Size\_mm\_min (Numeric), Granule\_Size\_mm\_max (Numeric)
* **Rationale and Data:** The size of fertilizer granules directly influences their surface-area-to-volume ratio, which in turn affects the rate at which they dissolve in soil moisture. More importantly, a consistent and uniform granule size is essential for accurate and even application with modern spreading equipment. Products with a narrow and specified size range, such as 2-4 mm or 1-6 mm, are indicative of higher manufacturing quality and will result in a more uniform distribution of nutrients across a field.8 This prevents "striping" in the field, where some areas are over-fertilized and others are under-fertilized. These features provide a subtle but powerful signal of product quality and predictable performance.

### **3.3. Advanced Nutrient Profile**

The simple N-P-K percentage is a blunt instrument. A deeper understanding requires dissecting the chemical forms of these nutrients and accounting for the presence of other essential elements.

#### **3.3.1. Nitrogen Forms**

The total nitrogen percentage tells only part of the story. The chemical form of that nitrogen dictates its release speed, its interaction with the soil, and its susceptibility to various loss pathways.

* **New Features:** N\_Form\_Urea\_pct (Numeric), N\_Form\_Ammoniacal\_pct (Numeric), N\_Form\_Nitrate\_pct (Numeric)
* **Rationale and Data:** These three forms of nitrogen have distinct agronomic characteristics:
  + **Urea Nitrogen (CO(NH2​)2​):** Before plants can use it, urea-N must be converted to ammoniacal-N by the urease enzyme in the soil. This conversion step makes it a slightly slower-release form.22 If urea is left on the soil surface, especially in warm and moist conditions, a portion of the nitrogen can be lost to the atmosphere as ammonia gas through a process called volatilization. This is a major pathway for nitrogen inefficiency.19 Many common fertilizers like Urea (46-0-0), 28-28-0, and balanced NPKs like 17-17-17 and 20-20-20 contain a significant portion of their nitrogen in the urea form.2
  + **Ammoniacal Nitrogen (NH4+​):** This form is positively charged and thus binds to negatively charged soil colloids (clay and organic matter). This binding makes it resistant to leaching, keeping it in the root zone.23 However, when soil microbes convert ammoniacal-N to nitrate-N (a process called nitrification), hydrogen ions (  
    H+) are released, which increases soil acidity.6 This acidifying effect can lower soil pH over time, potentially impacting the availability of other nutrients. Fertilizers such as DAP (18-46-0), 28-28-0, and 10-26-26 are rich in ammoniacal nitrogen.2
  + **Nitrate Nitrogen (NO3−​):** This form is immediately available for plant uptake without needing any conversion. However, it is negatively charged and does not bind to soil colloids, making it highly mobile in soil water and very susceptible to leaching below the root zone, especially during heavy rainfall or irrigation events.23 Some high-performance soluble fertilizers, like certain 20-20-20 blends, contain a portion of nitrate-N for rapid plant response.13

By deconstructing the total Nitrogen\_Percentage into these constituent forms, we encode a multi-dimensional profile of behavior into the dataset. A fertilizer like Gromor 28-28-0, which is specified to contain both urea (19%) and ammoniacal (9%) nitrogen, presents a complex profile.2 An XGBoost model can learn to use the relative percentages of these forms to predict its suitability under different scenarios. For example, it could learn that the high

N\_Form\_Urea\_pct makes it risky to apply without incorporation in warm weather (volatilization risk), while the N\_Form\_Ammoniacal\_pct contributes to a long-term acidifying effect that must be managed. This represents a far more sophisticated understanding than is possible from a simple '28' value for nitrogen.

#### **3.3.2. Secondary and Micronutrient Composition**

Modern, high-yield agriculture increasingly recognizes that crop performance is often limited by nutrients beyond just N, P, and K. The presence of secondary and micronutrients is a key differentiator between basic commodity fertilizers and premium, performance-oriented products.

* **New Features:** Has\_Sulphur (Boolean), Has\_Calcium (Boolean), Has\_Magnesium (Boolean), Has\_Zinc (Boolean), Has\_Boron (Boolean), Has\_Iron (Boolean), Has\_Manganese (Boolean), Has\_Copper (Boolean), Has\_Molybdenum (Boolean), Contains\_Micronutrient\_Package (Boolean)
* **Rationale and Data:** Secondary nutrients like Sulphur (S), Calcium (Ca), and Magnesium (Mg), and micronutrients like Zinc (Zn), Boron (B), and Iron (Fe) are essential for various plant metabolic functions. Deficiencies in these elements can severely limit growth and yield, even when NPK levels are adequate. Many advanced fertilizers are now formulated to include these vital elements. For example, NPK(S) 10:26:26(2) explicitly contains 2% Sulphur, making it highly suitable for crops with a high sulphur demand like oilseeds.20 Other products, such as certain 28-28-28 or 20-20-20 blends, are marketed with a "TE" (Trace Elements) package, indicating the presence of a suite of micronutrients.13 Conversely, the long-term use of some basic fertilizers like DAP can induce deficiencies of micronutrients like zinc and iron.28 Creating boolean flags for the presence of these nutrients allows the model to identify these premium formulations and correlate their use with specific crops or soil conditions where such nutrients are known to be beneficial.

## **Augmentation with Agronomic and Environmental Features**

Beyond the intrinsic properties of a fertilizer, its value is determined by how it behaves in the field. This section details features engineered to capture the dynamic aspects of nutrient release, soil interaction, and potential risks.

### **4.1. Nutrient Release Profile**

The timing of nutrient availability is as important as the quantity. The concept of "release rate" captures this temporal dimension.

* **New Feature:** Release\_Rate (Categorical: Fast, Slow, Controlled)
* **Rationale and Data:** This feature is a direct proxy for Nutrient Use Efficiency (NUE)—the proportion of applied nutrients that are actually taken up by the crop.
  + **Fast:** These fertilizers are typically highly water-soluble and make their nutrient load available almost immediately upon application. This is ideal for giving crops a quick boost, correcting deficiencies, or for use on short-season crops where rapid growth is required.14 Most standard granular NPKs, liquids, and powders fall into this category.
  + **Slow:** These fertilizers release nutrients over a more extended period. This category includes organic fertilizers, which rely on microbial decomposition, and products like Urea, which must first undergo enzymatic conversion in the soil before its nitrogen becomes fully available.14 The slower release helps reduce the risk of nutrient loss and provides a more sustained feeding for the crop.
  + **Controlled:** This refers to advanced formulations, often called Enhanced Efficiency Fertilizers (EEFs). These products use technologies like polymer coatings or specialized chemical structures (e.g., Triazone nitrogen) to release nutrients at a predictable rate over a specified period, such as 8-12 weeks.3 This technology is designed to perfectly synchronize nutrient supply with the crop's demand curve, maximizing NUE and minimizing environmental losses.30

The importance of this feature cannot be overstated. Applying a "Fast" release fertilizer long before the crop's period of peak nutrient demand (e.g., stem elongation in wheat or the V6-V8 stage in maize) is highly inefficient.31 The readily available nutrients, particularly nitrogen, are likely to be lost to the environment via leaching or volatilization before the plant can use them. A model equipped with the

Release\_Rate feature can learn that for long-duration crops or in environments prone to nutrient loss (e.g., sandy soils, high rainfall regions), a 'Slow' or 'Controlled' release rate is a powerful predictor of successful outcomes.

### **4.2. Soil Interaction Profile**

When a fertilizer is applied, it initiates a series of chemical and biological reactions with the soil. These interactions can have profound effects on the soil environment and nutrient availability.

#### **4.2.1. Impact on Soil pH**

The application of nitrogen fertilizer is a primary driver of changes in soil pH in agricultural systems.

* **New Feature:** Soil\_pH\_Effect (Categorical: Acidifying, Neutral, Alkalizing)
* **Rationale and Data:** A fertilizer's long-term effect on soil pH is one of its most critical attributes.
  + **Acidifying:** Fertilizers containing ammoniacal nitrogen (NH4+​), or those that produce it (like Urea), have an acidifying effect. This occurs when soil microbes convert the ammonium to nitrate, releasing hydrogen ions (H+) in the process.6 This is a significant effect for fertilizers like DAP, Ammonium Sulfate, and Urea.33 Over time, this acidification can lower the availability of nutrients like Phosphorus and Molybdenum and increase the concentration of toxic elements like Aluminum.6
  + **Neutral:** Some fertilizers are formulated to be pH-neutral, having minimal long-term impact on soil acidity. This is a desirable characteristic, especially for soils that are already in the optimal pH range.9
  + **Alkalizing:** Fertilizers containing nitrogen in the nitrate form (NO3−​) can have a slight alkalizing effect, as the plant consumes a hydrogen ion (or releases a hydroxide ion) when it takes up the nitrate molecule.24 DAP, while containing acidifying ammonium, has a temporarily alkaline reaction when first dissolved in water, which can be beneficial in acidic soils.28

This feature is a cornerstone for modeling nutrient availability. A model can learn that a fertilizer with an 'Acidifying' effect may be less effective on an already acidic soil unless a liming program is also in place.

#### **4.2.2. Environmental and Safety Risk Profile**

Effective fertilizer management involves balancing crop nutrition with the mitigation of potential negative impacts. These derived features synthesize multiple properties into actionable risk assessments.

* **New Features:** Leaching\_Potential (Categorical: High, Medium, Low), Phytotoxicity\_Risk (Categorical: High, Medium, Low)
* **Rationale and Data:** These features are not typically stated on a product label but are inferred through an expert assessment of the fertilizer's composition and known behavior.
  + **Leaching Potential:** This quantifies the risk of nutrients being washed below the plant root zone by water. The risk is primarily associated with highly mobile, water-soluble nutrients.
    - **High:** Attributed to fertilizers containing nitrate-nitrogen (NO3−​), which is highly soluble and not held by soil particles.24
    - **Medium:** Attributed to urea-based fertilizers. While urea itself is mobile, its conversion to ammonium can reduce leaching, but losses can still occur before this happens.19
    - **Low:** Attributed to ammoniacal-nitrogen (NH4+​), which binds to soil colloids, and phosphorus, which is relatively immobile in most soils.23 Controlled-release products are, by design, low-leaching.30
  + **Phytotoxicity Risk:** This is the potential for the fertilizer to cause "burn" or injury to seeds, seedlings, or plant roots. The risk stems from two main sources: high localized concentrations of salts (osmotic stress) and the production of toxic compounds like free ammonia (NH3​).38
    - **High:** Attributed to fertilizers with a high salt index or those containing urea, which can rapidly hydrolyze to produce damaging concentrations of ammonia, especially when placed too close to seeds.40 DAP also carries a risk due to ammonia release.33
    - **Medium:** Attributed to most standard soluble NPK fertilizers when used at recommended rates.
    - **Low:** Attributed to slow-release and controlled-release formulations, as well as organic fertilizers, which release nutrients gradually and avoid sharp spikes in salt concentration.14

The creation of these risk features is a powerful example of knowledge synthesis. No single source document might state, "Fertilizer X has a high phytotoxicity risk." Instead, this conclusion is derived by combining knowledge of its composition (e.g., high N\_Form\_Urea\_pct) with an understanding of the chemical processes that cause phytotoxicity (e.g., ammonia production from urea hydrolysis 40). This transforms multiple, disparate data points into a single, highly predictive feature that can inform recommendations on safe application methods (e.g., a 'High' risk fertilizer should not be placed directly in the seed furrow).

## **Augmentation with Crop-Specific Application Features**

While the preceding features describe universal properties, the optimal use of any fertilizer is ultimately context-dependent, varying significantly by crop. To capture this tactical knowledge, a set of crop-specific features was engineered for each major crop listed in the Benefit\_Crops column. This allows for the development of highly specialized models tailored to individual cropping systems.

### **Example for Maize (Zea mays)**

* **New Features:** Maize\_Rec\_Timing (Text), Maize\_Rec\_Method (Text), Maize\_Optimal\_Temp\_F\_min (Numeric), Maize\_Optimal\_Temp\_F\_max (Numeric)
* **Rationale and Data:**
  + **Timing:** The timing of nitrogen application is critical for maize. While a starter fertilizer is beneficial at planting, the peak demand for nitrogen occurs later. Applying the bulk of nitrogen top-dressing at the V6-V8 growth stage (when plants are approximately 6-8 inches tall) is ideal for maximizing uptake and yield.41 Applications made too early (e.g., 3-4 leaf stage) are inefficient, as the nitrogen is vulnerable to leaching before the crop's rapid uptake phase begins.31 For specific fertilizers like DAP, application at sowing is recommended to supply early phosphorus 5, whereas for Urea, split applications are often most effective.43
  + **Method:** Top-dressing can be done by broadcasting granules, though this carries a risk of lodging in the leaf whorl and causing foliar burn.31 More precise methods like banding or injection next to the row are more efficient. Foliar application is also an option, particularly for corrective treatments.31
  + **Environmental Conditions:** Maize is a warm-season crop, with optimal growth occurring in a temperature range of 75°F to 86°F.44 Growth ceases below 50°F. This temperature range is a critical feature for developing location-aware recommendation systems.

### **Application Profiles for Other Key Crops**

This process of extracting tactical information was repeated for the other key crops identified in the dataset:

* **Sugarcane:** A tropical grass that thrives in warm, humid climates with well-drained, loamy soils and a pH of 6.0-6.5.45 Its fertilization strategy is phased: fertilizers high in phosphorus (like 14-35-14) are best applied in the furrow at planting to stimulate root development. This is followed by top-dressing applications of nitrogen and potassium (e.g., using 17-17-17 or Urea) 8-10 weeks after planting to fuel vegetative growth.46
* **Wheat:** For winter wheat, phosphorus is critical for autumn root establishment. Therefore, DAP (18-46-0) is most effectively applied at or just before sowing.5 The majority of the crop's nitrogen requirement occurs in the spring. To maximize efficiency and minimize loss, the bulk of nitrogen fertilizer (like Urea or 20-20-0) should be top-dressed at spring green-up or during early stem elongation (Feekes Growth Stage 6).32
* **Cotton:** Cotton requires a split application strategy for nitrogen to match its developmental needs. An initial application should be made between planting and the 5-6 leaf stage, with subsequent applications during early squaring up to one week before bloom.48 To minimize volatilization losses, urea-based fertilizers should ideally be applied when soil temperatures are below 65°F and incorporated with irrigation or rainfall soon after.49
* **Groundnuts (Peanuts):** This legume benefits from fertilizers high in phosphorus and potassium, especially during the reproductive phases. Formulations like 10-26-26 are ideal for promoting flowering and pod development.50 Planting should occur when the soil is warm and moist to ensure good germination and establishment.51
* **Oilseeds (e.g., Rapeseed/Canola):** These crops have a high demand for both sulphur and nitrogen. A particularly critical and high-value intervention is the application of foliar nitrogen (often as foliar urea) towards the end of the flowering period. This late application keeps the canopy photosynthetically active for longer during the crucial seed-fill stage, directly contributing to higher yields and oil content.52

## **The Final Augmented Dataset and Data Dictionary**

The culmination of this feature engineering process is the augmented dataset, Fertilizer\_Augmented.csv, and its accompanying data dictionary. The dataset provides the structured information, while the dictionary provides the essential metadata, rationale, and transparency required for its effective use in a data science context.

### **Augmented Dataset: Fertilizer\_Augmented.csv**

The following table represents the final, enriched dataset. It incorporates the original columns alongside the newly engineered features, providing a comprehensive, multi-dimensional profile for each fertilizer.

| Fertilizer | Nitrogen\_Percentage | Phosphorous\_Percentage | Potassium\_Percentage | Benefit\_Soils | Benefit\_Crops | Physical\_Form | Solubility\_Class | Is\_Organic | Granule\_Size\_mm\_min | Granule\_Size\_mm\_max | N\_Form\_Urea\_pct | N\_Form\_Ammoniacal\_pct | N\_Form\_Nitrate\_pct | Has\_Sulphur | Contains\_Micronutrient\_Package | Release\_Rate | Soil\_pH\_Effect | Leaching\_Potential | Phytotoxicity\_Risk | Primary\_Application\_Method | Primary\_Application\_Timing |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 28-28 | 28 | 28 | 0 | Sandy,Black,Clayey,Loamy | Maize,Wheat,Barley,Cotton | Granular | High | FALSE | 1 | 6 | 19 | 9 | 0 | FALSE | FALSE | Slow | Acidifying | Medium | High | Broadcasting, Top-dressing | Basal, Top-dressing |
| 17-17-17 | 17 | 17 | 17 | Red,Sandy | Maize,Sugarcane,Cotton | Granular | High | FALSE | 2 | 4 | 14 | 3 | 0 | FALSE | FALSE | Fast | Acidifying | Medium | Medium | Broadcasting, Top-dressing | Initial application, Top-dressing |
| 10-26-26 | 10 | 26 | 26 | Red,Sandy | Ground Nuts,Pulses,Cotton | Granular | High | FALSE | 2 | 4 | 0 | 10 | 0 | TRUE | TRUE | Fast | Acidifying | Low | Low | Broadcasting, Banding | Basal (Autumn/Spring) |
| DAP | 18 | 46 | 0 | Red,Sandy | Maize,Wheat,Barley,Paddy,Cotton | Granular | High | FALSE | 1 | 4 | 0 | 18 | 0 | FALSE | FALSE | Fast | Acidifying | Low | High | Broadcasting, Banding | At Sowing |
| 20-20 | 20 | 20 | 0 | Red,Sandy | Maize,Wheat | Powder | High | FALSE | NA | NA | 20 | 0 | 0 | FALSE | FALSE | Fast | Acidifying | Medium | Medium | Foliar Spray, Fertigation | During growing season |
| 14-35-14 | 14 | 35 | 14 | Red,Black | Ground Nuts,Pulses,Cotton,Oil seeds | Granular | High | FALSE | 1 | 4 | 0 | 14 | 0 | TRUE | TRUE | Fast | Neutral | Low | Low | Broadcasting, Banding | Basal application |
| Urea | 46 | 0 | 0 | Red,Sandy | Maize,Sugarcane,Cotton | Granular | High | FALSE | 1 | 4 | 46 | 0 | 0 | FALSE | FALSE | Slow | Acidifying | Medium | High | Top-dressing, Incorporation | Split applications, Top-dressing |

### **Data Dictionary for Fertilizer\_Augmented.csv**

This dictionary is a critical component, providing the definitions, rationale, and source evidence for each new feature. It bridges the gap between agronomic science and data science, ensuring the user can understand, trust, and correctly interpret the features and the model's behavior.

| Feature Name | Data Type | Description and Rationale | Possible Values / Units | Source Snippets |
| --- | --- | --- | --- | --- |
| Physical\_Form | Categorical | The physical state of the fertilizer, which dictates application method and handling. Granular for broadcasting, Powder/Liquid/Gel for dissolving. | Granular, Powder, Liquid, Gel | 3 |
| Solubility\_Class | Categorical | The fertilizer's ability to dissolve in water. High solubility implies rapid nutrient availability but also higher leaching risk for mobile nutrients. | High, Medium, Low | 8 |
| Is\_Organic | Boolean | Indicates if the fertilizer is derived from natural, organic sources. Organic fertilizers typically have slower nutrient release and improve soil structure. | TRUE, FALSE | 11 |
| Granule\_Size\_mm\_min | Numeric | The minimum diameter of the fertilizer granules. A defined range indicates higher quality control and more uniform application. | millimeters (mm) | 8 |
| Granule\_Size\_mm\_max | Numeric | The maximum diameter of the fertilizer granules. A defined range indicates higher quality control and more uniform application. | millimeters (mm) | 20 |
| N\_Form\_Urea\_pct | Numeric | The percentage of total weight that is nitrogen in the urea form. Urea-N is prone to volatilization loss and requires enzymatic conversion. | % of total weight | 2 |
| N\_Form\_Ammoniacal\_pct | Numeric | The percentage of total weight that is nitrogen in the ammoniacal (NH4+​) form. This form is less prone to leaching but has an acidifying effect on soil. | % of total weight | 2 |
| N\_Form\_Nitrate\_pct | Numeric | The percentage of total weight that is nitrogen in the nitrate (NO3−​) form. This form is immediately available to plants but is highly prone to leaching. | % of total weight | 13 |
| Has\_Sulphur | Boolean | Indicates the presence of Sulphur (S), a critical secondary nutrient, especially for oilseeds and protein formation. | TRUE, FALSE | 20 |
| Contains\_Micronutrient\_Package | Boolean | Indicates if the fertilizer is fortified with a blend of trace elements (TE) such as Zn, B, Fe, Mn, etc. | TRUE, FALSE | 13 |
| Release\_Rate | Categorical | Describes the speed at which nutrients become available to the plant. This is a key factor for nutrient use efficiency. | Fast, Slow, Controlled | 14 |
| Soil\_pH\_Effect | Categorical | The long-term impact of the fertilizer on soil pH. Ammonium-based fertilizers are typically acidifying. | Acidifying, Neutral, Alkalizing | 6 |
| Leaching\_Potential | Categorical | An expert-assessed risk of mobile nutrients (primarily nitrate) being lost from the root zone due to water movement. | High, Medium, Low | 19 |
| Phytotoxicity\_Risk | Categorical | An expert-assessed risk of the fertilizer causing injury ("burn") to seeds or seedlings due to high salt concentration or ammonia toxicity. | High, Medium, Low | 38 |
| Primary\_Application\_Method | Text | The most common or recommended method for applying the fertilizer. | e.g., Broadcasting, Banding, Foliar Spray | 2 |
| Primary\_Application\_Timing | Text | The most common or recommended timing for application in a crop's life cycle. | e.g., Basal, At Sowing, Top-dressing | 2 |

## **Recommendations for XGBoost Model Implementation**

With the Fertilizer\_Augmented.csv dataset, a powerful and nuanced predictive model can be developed. The following recommendations are provided to guide the implementation process for an XGBoost model.

### **Data Preprocessing**

A robust preprocessing pipeline is essential to prepare the augmented data for the model.

* **Categorical Features:** Nominal categorical features with low cardinality, such as Physical\_Form, Release\_Rate, Solubility\_Class, and Soil\_pH\_Effect, should be converted into a numerical format using one-hot encoding. This creates a new binary column for each category, allowing the model to treat them independently without assuming an ordinal relationship.
* **Text Features:** The original Benefit\_Soils and Benefit\_Crops columns contain multiple labels within a single string. These should be processed using a multi-hot encoding (or binary encoding) approach. This involves creating a new binary feature for each unique soil type and crop type (e.g., Soil\_Sandy, Soil\_Clayey, Crop\_Maize, Crop\_Wheat). A given fertilizer's row will then have a '1' for each soil and crop it benefits, and a '0' otherwise. This transforms the unstructured text into a structured format that the model can effectively use.
* **Numerical Features:** While XGBoost is generally robust to the scale of input features, standardizing the numerical columns (e.g., Nitrogen\_Percentage, Granule\_Size\_mm\_max) using a method like StandardScaler is considered good practice. It ensures that all features are on a comparable scale, which can sometimes aid in faster convergence and prevent features with large magnitudes from dominating the model training process.
* **Handling Missing Values:** The Granule\_Size features are not applicable (NA) for non-granular forms. This should be handled appropriately, either by imputing a specific value (e.g., 0) or by using a model version that can handle missing values natively.

### **Feature Importance and Interpretation**

The augmented dataset is hypothesized to reveal a more insightful feature hierarchy than the original data.

* **Expected Feature Importance:** It is anticipated that the newly engineered features will rank highly in the feature importance plots generated by the trained XGBoost model. Specifically, features like Soil\_pH\_Effect, Release\_Rate, Leaching\_Potential, and the N\_Form\_\*\_pct columns are expected to demonstrate higher predictive power (higher F-score or gain) than the raw Nitrogen\_Percentage alone, as they encapsulate more complex behavioral information.
* **Interaction Analysis:** The true power of this approach will be revealed through the analysis of feature interactions. Tools like SHAP (SHapley Additive exPlanations) should be used to visualize and quantify these interactions. For example, one could plot the SHAP value for Phosphorous\_Percentage as a function of Soil\_pH\_Effect. The analysis would likely show that high phosphorus has a much stronger positive impact on the model's prediction (a higher SHAP value) when the Soil\_pH\_Effect is 'Neutral' compared to when it is 'Acidifying'. This would provide a visual, data-driven confirmation of the known chemical principle that phosphorus availability is pH-dependent.

### **Model Training Strategy**

* **Target Variable:** The user's query does not specify a target variable, but the augmented dataset is now versatile enough to support various predictive tasks. A potential target could be a categorical variable like Optimal\_Fertilizer\_for\_Crop\_X, or a continuous variable like Predicted\_Yield\_Response or Environmental\_Risk\_Score if such outcome data were available.
* **Cross-Validation:** A robust k-fold cross-validation strategy (e.g., 5 or 10 folds) must be employed during training. This ensures that the model's performance is evaluated on unseen data within the training set, providing a reliable estimate of its ability to generalize to new, external data and preventing overfitting.
* **Hyperparameter Tuning:** The default parameters of XGBoost may not be optimal for this specific dataset. A systematic hyperparameter tuning process, using techniques like GridSearchCV or RandomizedSearchCV, should be conducted. Key parameters to tune include n\_estimators (number of trees), max\_depth (complexity of each tree), learning\_rate (step size shrinkage), and gamma (minimum loss reduction to make a split), to find the combination that yields the best performance on the cross-validation sets.

#### Works cited

1. Fertilizer.csv
2. Gromor 28-28-0 – Coromandel: Fertiliser Company for Crop Protection & Nutrients, accessed June 29, 2025, <https://www.coromandel.biz/gromor-28-28-0/>
3. 1kg NPK 28-28-00 Water Soluble Foliar Fertilizer - Krushi Bio Care, accessed June 29, 2025, <https://www.krushibiocare.com/1kg-npk-28-28-00-water-soluble-foliar-fertilizer-8730028.html>
4. Diammonium Phosphate (DAP), accessed June 29, 2025, <https://www.toros.com.tr/documents/TARIMSAL/%C3%9Cr%C3%BCnler/Bitki%20Besleme/Klasik%20G%C3%BCbre-%C4%B0NG/DAP_Diamonyum_Fosfat_en.pdf>
5. Best Practices in Wheat Production - CIMMYT, accessed June 29, 2025, <https://www.cimmyt.org/content/uploads/2022/09/Zone-Specific-Recommendations-for-Central-Region.pdf>
6. The Role Of Nitrogen Fertilizer In Soil pH Levels - No-Till Farmer, accessed June 29, 2025, <https://www.no-tillfarmer.com/articles/2039-the-role-of-nitrogen-fertilizer-in-soil-ph-levels>
7. Soil pH Effects on Potassium and Phosphorus Fertilizer Availability and Management, accessed June 29, 2025, <https://nutrien-ekonomics.com/news/soil-ph-effects-potassium-and-phosphorus-fertilizer-availability-and-management/>
8. NPK (17-17-17) Fertilizer Application Features & Benefits, accessed June 29, 2025, <https://buoyancy.ae/wp-content/uploads/2023/12/NPK-17-17-17.pdf>
9. Compound High Quality Factory Direct Price Compound Organic Complex NPK Fertilizer for Crops, accessed June 29, 2025, <https://mingyuchemical.en.made-in-china.com/product/zwaAPWIHXORq/China-Compound-High-Quality-Factory-Direct-Price-Compound-Organic-Complex-NPK-Fertilizer-for-Crops.html>
10. Gromor 14-35-14 Fertilizer at Best Price in Secunderabad, Telangana | Coromandel International Limited - Tradeindia, accessed June 29, 2025, <https://www.tradeindia.com/products/gromor-14-35-14-fertilizer-2084895.html>
11. NPK-10-26-26 100% Water Soluble Mixture Fertilizer - gitaji life sciences private limited, accessed June 29, 2025, <https://www.gitajilifesciences.com/npk-10-26-26-100-water-soluble-mixture-fertilizer-9584505.html>
12. NPK 10:26:26 Liquid - Arihant Agro Care, accessed June 29, 2025, <https://arihantagrocare.in/en/products/npk-10-26-26-liquid>
13. 20-20-20 General Purpose Fertilizer - Nabteh.com, accessed June 29, 2025, <https://nabteh.com/products/20-20-20-general-purpose-fertilizer>
14. The Ultimate Guide to 20-20-20 Plant Fertilizer: A Deep Dive into Organic, Water Soluble Fertilizer Options - loyal chemical, accessed June 29, 2025, <https://loyalfertilizer.com/blog/organic-20-20-20-fertilizer/>
15. Sansar Green® NPK 14:35:14 for Plants(200 gm) Liberia | Ubuy, accessed June 29, 2025, <https://www.liberia.ubuy.com/productin/MBOQU8T7E-sansar-green-npk-14-35-14-for-plants-200-gm>
16. How to Use 20-20-20 Fertilizer: A Comprehensive Guide - Olimpum, accessed June 29, 2025, <https://olimpum.com/en/how-to-use-20-20-20-fertilizer-a-comprehensive-guide/>
17. Understanding Diammonium Phosphate: A Comprehensive Guide, accessed June 29, 2025, <https://loyalfertilizer.com/blog/understanding-diammonium-phosphate-a-comprehensive-guide/>
18. Water-soluble and Controlled-release Fertilization - Purdue Extension, accessed June 29, 2025, <https://www.extension.purdue.edu/extmedia/ho/ho-251-w.pdf>
19. How to Handle Urea - Successful Farming, accessed June 29, 2025, <https://www.agriculture.com/crops/fertilizers/how-to-handle-urea_174-ar2055>
20. Mineral fertilizers: liquid, nitrogen, nitrogen-phosphorus and complex - PhosAgro [EN], accessed June 29, 2025, <https://www.phosagro.com/production/fertilizer/azotno-fosfornye-udobreniya/diammofoska-npk-s-102626-2/>
21. Micronutrient NPK Fertilizer 10-26-26 Factory Price - Chemical, Boron | Made-in-China.com, accessed June 29, 2025, <https://mingyuchemical.en.made-in-china.com/product/hxrpqdRXXVYB/China-Micronutrient-NPK-Fertilizer-10-26-26-Factory-Price.html>
22. Urea Properties It is a granular, white coloured, highly soluble fertilizer and contains the highest nitrogen among all nitrogen, accessed June 29, 2025, <https://www.toros.com.tr/wp-content/uploads/2024/12/ure_en.pdf>
23. Nutrient Management :: Fertilizers - TNAU Agritech Portal, accessed June 29, 2025, <https://agritech.tnau.ac.in/agriculture/agri_nutrientmgt_fertilizers.html>
24. Fertilizers and Soil Acidity - Mosaic Crop Nutrition, accessed June 29, 2025, <https://www.cropnutrition.com/resource-library/fertilizers-and-soil-acidity/>
25. NPK 10-26-26 - OCP Brasil, accessed June 29, 2025, <https://www.ocpbrasil.com.br/sites/default/files/2023-04/OCP_Brasil_NPK_10-26-26_eng.pdf>
26. Gravity S 20-20-20 Micro - Heritage PPG, accessed June 29, 2025, <https://heritageppg.com/products/gravity-s-20-20-20>
27. NPK 28 28 28 - Gel Max Fertilizers Series - Grow Plant - Vaki-Chim, accessed June 29, 2025, <https://growplant.org/gel-fertilizers/max-28-28-28-te/>
28. Understanding Diammonium Phosphate (DAP): Composition, uses, and implications, accessed June 29, 2025, <https://agrimachinery.africa/nutrients-application/understanding-diammonium-phosphate-dap-composition-uses-and-implications/>
29. 28-0-0, accessed June 29, 2025, <https://assets.greenbook.net/22-58-03-11-10-2023-Drexel_X_28-0-0_-_label.pdf>
30. Fertilizer Guide - Fast Growing Trees, accessed June 29, 2025, <https://www.fast-growing-trees.com/pages/fertilizers>
31. Top Dressing Fertilizer For Maize: Why Timing is Key, accessed June 29, 2025, <http://cereals.co.ke/index.php/featured-articles/231-top-dressing-fertilizer-for-maize-why-timing-is-key>
32. When is the Best Time to Apply N to Wheat | Agronomic Crops Network, accessed June 29, 2025, <https://agcrops.osu.edu/newsletter/corn-newsletter/2016-05/when-best-time-apply-n-wheat>
33. DAP Fertilizer Side Effects – What Every HSR Layout Gardener Should Wa - Idyl, accessed June 29, 2025, <https://www.idyl.co.in/blogs/blog/dap-fertilizer-side-effects-what-every-hsr-layout-gardener-should-watch-for>
34. www2.gov.bc.ca, accessed June 29, 2025, <https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/agriculture-and-seafood/agricultural-land-and-environment/soil-nutrients/600-series/631001-2_nitrogen_fertilization_as_related_to_soil_ph.pdf>
35. GROMOR 14-35-14 - Bharat Enterprises, accessed June 29, 2025, <https://bharat-enterprises.co.in/gromor-14-35-14/>
36. Yarafert's Role in Reducing Runoff and Nutrient Leaching, accessed June 29, 2025, <https://yarafert.com/yara-runoff-reduction/>
37. Are phosphate leaching losses in soil considerable? - ResearchGate, accessed June 29, 2025, <https://www.researchgate.net/post/Are-phosphate-leaching-losses-in-soil-considerable>
38. KTS 0-0-25, 17% Sulfur, Potassium Thiosulfate | Plant Food Company, Inc., accessed June 29, 2025, <https://www.plantfoodco.com/specialty-agriculture/products/liquid-fertilizer/0-0-25-17-sulfur/>
39. Phytotoxicity: Definitions, Causes, Symptoms and Solutions - Dora Agri-Tech, accessed June 29, 2025, <https://doraagri.com/phytotoxicity-definitions-causes-symptoms-and-solutions/>
40. Toxicity as a cause of the inefficiency of urea as a fertilizer. Part I - ResearchGate, accessed June 29, 2025, <https://www.researchgate.net/publication/230300680_Toxicity_as_a_cause_of_the_inefficiency_of_urea_as_a_fertilizer_Part_I>
41. Timing is Everything: Top-Dressing Fertilizer in Maize - YouTube, accessed June 29, 2025, <https://www.youtube.com/watch?v=Q_QroJIFdDY>
42. The ideal time and method for DAP fertilizer application in wheat crop | Abid Ali Agrarian, accessed June 29, 2025, <https://www.youtube.com/watch?v=BU0yXpYzAnE>
43. What was the optimal application depth of urea to improve the nitrogen use efficiency in summer maize? | ResearchGate, accessed June 29, 2025, <https://www.researchgate.net/post/What-was-the-optimal-application-depth-of-urea-to-improve-the-nitrogen-use-efficiency-in-summer-maize>
44. Sweet Corn Production - Purdue Extension, accessed June 29, 2025, <https://www.extension.purdue.edu/extmedia/NCH/NCH-43.html>
45. Fertilizers for Sugarcane - Crops - ICL Growing Solutions, accessed June 29, 2025, <https://icl-growingsolutions.com/en-us/agriculture/crops/growing-sugarcane-crop-nutrition-advice/>
46. Fertilize Correctly - SIA, accessed June 29, 2025, <https://www.jamaicasugar.org/fertilize-correctly.html>
47. Fertilizing Winter Wheat | NDSU Agriculture, accessed June 29, 2025, <https://www.ndsu.edu/agriculture/extension/publications/fertilizing-winter-wheat>
48. Understanding Cotton Growth: Fertility, Nutrition, and Nutrient Management, accessed June 29, 2025, <https://www.cottongrower.com/cotton-production/understanding-cotton-growth-fertility-nutrition-and-nutrient-management/>
49. Urea-Based Fertilizers in Forage Production | Mississippi State University Extension Service, accessed June 29, 2025, <http://extension.msstate.edu/publications/publications/urea-based-fertilizers-forage-production>
50. Understanding the Applications of NPK 10-26-26 Fertilizer for Optimal Plant Growth, accessed June 29, 2025, <https://www.hhfertilizer.com/news/npk102626uses-11150.html>
51. 4R-Groundnut-Guide-N-Ghana.pdf - African Plant Nutrition Institute (APNI), accessed June 29, 2025, <https://www.apni.net/wp-content/uploads/2023/06/4R-Groundnut-Guide-N-Ghana.pdf>
52. Agronomy Advice | Increase your oilseed yields with foliar nitrogen - Yara UK, accessed June 29, 2025, <https://www.yara.co.uk/crop-nutrition/agronomy-advice/oilseed-foliar-urea/>