**Table 1: Data curation process for potato traits, diseases and genetic information**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **There is a potato variety called** | Accent | Accent is a high yielding early bulking table variety. Attractive, uniform oval tubers with light yellow skins and flesh and excellent cooking characteristics. Resistance to common scab, bruising and splitting. Resistant to potato cyst nematode *Globodera rostochiensis*i Ro1. Susceptible to foliage and tuber late blight, blackleg and potato cyst nematode *Globodera pallida* Pa 2/3 and 1. Sensitive to Sencorex post emergence. | **this variety is as a result of a genetic cross between** | Alcmaria and AM66.42 | **and it was developed in** | Netherlands | **. This variety was bred at approximately the year** | undetermined | **. The shape of this variety is** | Short - oval | **and the colour of the skin is** | Cream | **. The colour of its flesh is** | Light yellow | **and the depth of its eye is** | Shallow - medium | **. The smoothness of its skin is** | Medium | **while the colour of its base lightsprout is** | Pink | . **It is a** | First Early | **maturing variety** | **with a** | Short | **height** | **while the flower colour is** | White | **as its frequency of berries is** | Absent | **. On a scale of 1 to 9 its reaction to various diseases and disorders are as follows with 1 to 3 being resistant and 4 to 6 being tolerant and 7 to 9 being susceptible** | **to bruising is** | 8 | **, to splitting is** | 7 | **, to Potato Virus YO is** | 4 | **, to Potato Virus YN is** | unknown | **, to Potato Virus A is** | unknown | , **to Potato mop top (spraing) is** | unknown | **, to Potato Leafroll Virus is** | 4 | **, to Potato Cyst Nematode (Pa 2/3 - 1) (Globodera pallida Pa 2/3 1) is** | 2 | **, to Potato Cyst Nematode (Ro1) (Globodera rostochiensis Ro1) is** | 9 | **, to common scab (Streptomyces scabiei) is** | 7 | **, to blackleg (Pectobacterium atrosepticum) is** | 2 | **, to dry rot (Fusarium sulphureum) is** | unknown | **, to dry rot (Fusarium coeruleum) is** | unknown | **, to skin spot (Polyscytalum pustulans) is** | unknown | **, to silver scurf (Helminthosporium solani) is** | unknown | **, to black dot (Colletotrichum coccodes) is**  unknown  **, to powdery scab (Spongospora subterranea) is**  6  **, to late blight on tubers (Phytophthora infestans) is**  2  **and last but not least, to late blight on foliage (Phytophthora infestans) is**  3  **. The variety has a**  undetermined  **dormancy period**  a**nd a dry matter of**  undetermined percentage  **this variety has a**  undetermined  **determinacy**  **. It is a**  uncharacterized  **cooking type** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **There is a potato variety called** | Jinshu No.16 | **with an international potato center (CIP) identification number** |  | **it was developed in** | China | **in the year** | unknown | **as a cross between** | NL94014 | **and** | 9333-11 | **and released in the year** | 2007 | **by the research institution of** | High Latitude Crop Research Institute, Shanxi Academy of Agricultural Sciences | **as a** | NARS crosses from CIP progenitors |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **There is a potato variety called** | Maria Huanca | **with an international potato center (CIP) identification number** | CIP279142.12 | **it was developed in** | Peru | **in the year** | 1979 | **as a cross between** | AM 66-426 | **and** | 276012.24 | **and released in the year** | 1987 | **by the research institution of** | **which is unknown** | **as a** | NARS selections from CIP crosses |

Tabular representation for curated concatenated tabular to text sentence conversion. The cells in bold are the curated joining statements while the non bold are the original tabular column data

This data was then structured for large language model training as below with an instruction, premise, hypothesis and label column

**Table 2: Data structure for training the LLM transformer models**

|  |  |  |  |
| --- | --- | --- | --- |
| **instructions** | **premise** | **hypothesis** | **label** |
| A description of a potato variety Accent including its market and agronomic traits relative to its disease resistances | There is a potato variety called Accent Accent is a high yielding early bulking table variety. Attractive, uniform oval tubers with light yellow skins and flesh and excellent cooking characteristics. Resistance to common scab, bruising and splitting. Resistant to potato cyst nematode Globodera rostochiensisi Ro1. Susceptible to foliage and tuber late blight, blackleg and potato cyst nematode Globodera pallida Pa 2/3 and 1. Sensitive to Sencorex post emergence. this variety is as a result of a genetic cross between Alcmaria and AM66.42 and it was developed in Netherlands . This variety was bred at approximately the year undetermined . The shape of this variety is Short - oval and the colour of the skin is Cream . The colour of its flesh is Light yellow and the depth of its eye is Shallow - medium . The smoothness of its skin is Medium while the colour of its base lightsprout is Pink . It is a First Early maturing variety with a Short height while the flower colour is White as its frequency of berries is Absent . On a scale of 1 to 9 its reaction to various diseases and disorders are as follows with 1 to 3 being resistant and 4 to 6 being tolerant and 7 to 9 being susceptible to bruising is 8 , to splitting is 7 , to Potato Virus YO is 4 , to Potato Virus YN is unknown , to Potato Virus A is unknown , to Potato mop top (spraing) is unknown , to Potato Leafroll Virus is 4 , to Potato Cyst Nematode (Pa 2/3 - 1) (Globodera pallida Pa 2/3 – 1) is 2 , to Potato Cyst Nematode (Ro1) (Globodera rostochiensis Ro1) is 9 , to common scab (Streptomyces scabiei) is 7 , to blackleg (Pectobacterium atrosepticum) is 2 , to dry rot (Fusarium sulphureum) is unknown , to dry rot (Fusarium coeruleum) is unknown , to skin spot (Polyscytalum pustulans) is unknown , to silver scurf (Helminthosporium solani) is unknown , to black dot (Colletotrichum coccodes) is unknown , to powdery scab (Spongospora subterranea) is 6 , to late blight on tubers (Phytophthora infestans) is 2 and last but not least, to late blight on foliage (Phytophthora infestans) is 3 . The variety has a undetermined dormancy period and a dry matter of undetermined percentage this variety has a undetermined determinacy . It is a uncharacterized cooking type | 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture | 1 |
| A description of a potato variety Jinshu No.16 including its market and agronomic traits relative to its disease resistances | There is a potato variety called Jinshu No.16 with an international potato center (CIP) identification number it was developed in China in the year unknown as a cross between NL94014 and 9333-11 and released in the year 2006 by the research institution of High Latitude Crop Research Institute, Shanxi Academy of Agricultural Sciences as a NARS crosses from CIP progenitors | 2.5 By 2020, maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species, including through soundly Transforming our world: the 2030 Agenda for Sustainable Development managed and diversified seed and plant banks at the national, regional and international levels, and promote access to and fair and equitable sharing of benefits arising from the utilization of genetic resources and associated traditional knowledge, as internationally agreed | 1 |
| A description of a potato variety Maria Huanca including its market and agronomic traits relative to its disease resistances | There is a potato variety called Maria Huanca with an international potato center (CIP) identification number CIP279142.12 it was developed in Peru in the year 1979 as a cross between AM 66-426 and 276012.24 and released in the year 1987 by the research institution of which is unknown as a NARS selections from CIP crosses | 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture | 1 |
| This is a description of plant pathogen that causes plant disease and the integrated pest management strategy herein | Late Blight of Potato Symptoms After blossoming, large, dark green, water-soaked spots appear on leaves in wet Late Blight on Potato weather, first on lower leaves. As a spot enlarges the center is shriveled, dry, dark brown to black, and a downy, whitish growth appears on the underside of leaves. Similar lesions are formed on stems and petioles, and there is a characteristic strong odor as tops are blighted. On tubers, first symptoms are small brown to purple discolorations of skin on upper side, changing to depressed pits when tubers are removed from soil and put in storage (4). On cutting through the potato, a reddish brown dry rot is seen. Life History The primary cycle starts with infected tubers, which have harbored mycelium in the dry rot patches over winter. If infected seed pieces are planted, the fungus grows systemically into the shoots and finally fruits by sending sporangiophores out through the stomata on lower leaf surfaces (5). These swell at the tips into ovoid bodies, sporangia, then branch and produce successively more sporangia. The latter may function as conidia, putting out a germ tube, but more often are differentiated into a number of swarmspores (zoospores), which have cilia enabling them to swim about after they are splashed by rain to another leaf. Eventually they stop swimming and send a germ tube in through the leaf cuticle or enter through a stoma. Initial infection in the field also comes from conidia blown over from sprouts produced on infected tubers in cull piles. Blighting follows rapidly, with first symptoms 5 days or less from the time of infection and with the fungus fruiting again in a whitish layer on the underside of leaves. Late Blight of Potatoes. Sporangiophores of Phytophthora infestans emerging from leaf, bearing sporangia, sometimes called conidia, which germinate by zoospores Tubers with only a thin covering of soil may be infected by swarmspores washing down onto them from blighted leaves overhead; they are also infected during digging if it is done in moist weather while tops are still green. Swarmspores remain viable in the soil several weeks while awaiting favorable conditions. Oospores, the sexual spores, are apparently not required in the life cycle for they are not found with potatoes grown in the field. They have been produced in culture. Weather Relations This is a disease entirely dependent on weather conditions. Temperature and moisture conditions are right for an epiphytotic about 2 years out of 5. Zoospores are produced only in cool weather, 60 F and under, but they invade leaves most rapidly at higher temperatures. Because they are swimming spores, rain is required. A cool, wet July is usually followed by blight in August and September. Control Some varieties, such as Kennebec, Essex, Pungo, and Cherokee, are resistant to the common strain of the fungus but not to some of the newer strains. Treat potato dumps and cull piles with a weed spray to control sprouts. Delay digging crop until 2 weeks after tops die, or else kill the tops with a weed killer to prevent infection at early digging. Late Blight of Tomato Although there are potato and tomato strains of Phytophthora infestans, each is capable of infecting the other host. Ordinarily blight starts with potatoes in midsummer; when the fungus moves over to tomatoes, it has to go through several cycles to build up a strain virulent enough to produce general blighting, and by that time the tomato season is nearly over. Now we know that it is possible for the tomato strain to winter in potato tubers and be ready to inflict damage on tomatoes with the first crop of zoospores produced on potato sprouts. Conversely, tomato seedlings brought up from the South and planted near potato fields can start an epiphytotic of late blight on potatoes. | 9.4 By 2030, upgrade infrastructure and retrofit industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes, with all countries taking action in accordance with their respective capabilities | 0 |

curated\_agric\_dataset\_v5.csv Data used to train both transfoemer model llama3.2 and MPnet sentence transformer

DATA.json Data store from the google sheet weather data. Data can be used when the model is offline and updates when online

potatoDiseaseprofile.csv Potato Disease data white ideal ranges for temp humidity and soil ph

SDGTargets.txt SDG targets used in the training

soilData.csv Soil data including soil pH data contains for Kenya only the others can be updated once available

EpidemiologyAgroecologicalZones.csv Data for cropland zones

QA\_data.csv question and answer data for training larger model in agriculture. Data not used here because of the limited training resources